DOCUMENT RESUME

ED 056 647 24 HE 002 646

AUTHOR	Ronan, W. W.
TITLE	Development of an Instrument to Evaluate College Classroom Teaching Effectiveness. Final Report.
INSTITUTION	Georgia Inst. of Tech., Atlanta.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-1-D-045
PUB DATE	13 Sep 71
GRANT	OEG-4-71-0067
NOTE	169p.
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	College Students; College Teachers; *Effective Teaching; *Evaluation; Faculty; *Higher Education; *Taxonomy; Teacher Behavior; *Teacher Evaluation

ABSTRACT

This research study was intended as a basis for developing a device to measure college classroom teaching effectiveness. A review of some of the literature indicated that most past efforts to evaluate college teaching had used some sort of rating form and the results were quite ambiguous. However, 3 studies using the critical incident technique showed some promise for developing an evaluation device and that technique was used for this study. Students at Georgia Institute of Technology collected some 3,000 incidents describing "best" and "worst" teachers. These incidents were categorized into behavioral areas describing the dimensions of effective and ineffective teaching. The results were quite similar to those from other critical incident studies. The general conclusion was that a taxonomy of teaching behaviors has been isolated and that these behaviors can be described using the behavioral statements contained in the body and appendices of this report. The behavioral statements can be used to describe effective and ineffective teaching behaviors as seen by students. Further research is needed to establish the psychometric characteristics of the behavioral items. (Author/AF)

ED0 56647

OE-NCERD
HE

FINAL REPORT

Project No. 1-D-045
Grant No. OEG-4-71-0067

W. W. RONAN

GEORGIA INSTITUTE OF TECHNOLOGY

Atlanta, Georgia

DEVELOPMENT OF AN INSTRUMENT TO EVALUATE COLLEGE CLASSROOM TEACHING EFFECTIVENESS

September 13, 1971

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

HE002646

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education

National Center for Educational Research and Development
(Regional Research Program)

FINAL REPORT

Project No. 1-D-045
Grant No. OEG-4-71-0067

DEVELOPMENT OF AN INSTRUMENT TO EVALUATE COLLEGE CLASSROOM TEACHING EFFECTIVENESS

W. W. RONAN

GEORGIA INSTITUTE OF TECHNOLOGY

Atlanta, Georgia

September 13, 1971

The research reported herein was performed pursuant to a grant with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
National Center of Educational Research and Development

Errata

Page 26	Third paragraph from top, tenth line read results for rsults
Page 33	Third paragraph from bottom, second line read not for now

ABSTRACT

This research study was intended as a basis for developing a device to measure college classroom teaching effectiveness. A review of some of the relevant literature indicated that most past efforts to evaluate college teaching had used some sort of rating form and the results were quite ambiguous. However, three studies using the critical incident technique showed some promise for developing an evaluation device and the technique was used here.

Students at Georgia Institute of Technology collected some 3,000 incidents describing "best" and "worst" teachers. These incidents were categorized into behavioral areas describing the dimensions of effective and ineffective teaching. The results were quite similar to those from other critical incident studies.

The general conclusion of the study is that a taxonomy of teaching behaviors has been isolated and that these behaviors can be described using the behavioral statements contained in the body and appendices of this report. The behavioral statements can be used to describe effective and ineffective teacher behaviors as seen by students. Further research is needed to establish the psychometric characteristics of the behavioral items.

PREFACE

The recent heightened interest in the consequences of higher education has resulted in some innovation but relatively little basic change. Colleges and universities remain much as they were a decade ago prior to demands for change. However, there is some indication that the parties-at-interest are becoming aware that the future of institutions of higher education is being decided now. One point that all parties agree on, albeit with somewhat shaky evidence, is that the quality of higher education leaves something to be desired. This fact coupled with the changing values of our society and changing conditions pertaining to the role of college in the lives of young people have led to near crisis on many college compuses. Where college once provided an escape from the draft, the entry to an assured and high paying occupation, and a desirable way to consume four or more years, the trend today is much different. Students, and many faculty, today demand much more intrinsic worth of the college experience and one of the principle determinants of intrinsic worth is the quality of instruction. Students are less likely to tolerate incompetence today, they are openly critical of absenteeism, grantsmanship, narrow provincialisms and the disdain of professors. Clearly, college faculty and institutions as a whole are being taken to task, are being held accountable for their contribution to student accomplishment of academic and personal goals. The result is that the emphasis on quality instruction, innovation in methods and institutional characteristics is increasing. Unfortunately, progress has been slow in coming and it is generally agreed that the lack of suitable criteria for evaluating teacher performance has been the major block to such progress. This research was conceived as a step toward developing such criteria.

The author would like to express his appreciation to Mr. Gary Latham, former graduate student at Georgia Tech, for his original classification of the research data used in this study. To Mr. James Boileau, a senior student, for doing the second classification, and to Dr. Erich Prien, Memphis State University, and to Dr. Edmond Marks, Pennsylvania State University, for reading and commenting on the manuscript.

Miss Deborah Ehrenfeld served most capably as Research Assistant for the project.

The author is fully responsible for all phases of the final report.

TABLE OF CONTENTS

Page

ABSTRACT i

PREFACE ii

INTRODUCTION 1

SOME RELEVANT RESEARCH 3

- Ratings of Teacher Performance 10
- Factor Analytic Studies of Teaching Behavior 29
- Critical Incident Studies of Teaching Performance 37

RESEARCH PROCEDURE 42

RESEARCH RESULTS 46

- Effective Behaviors 46
- Ineffective Behaviors 73
- Dimensions of Effective and Ineffective Teaching Performance 107
- Evaluative Questions - Effective and Ineffective Teaching Performance 119

SUMMARY AND CONCLUSIONS 124

BIBLIOGRAPHY 128

APPENDICES 134

6

LIST OF TABLES

		Page
TABLE 1	Distribution of Student Ratings for Selected Questions	13
TABLE 2	Reported Effective Incidents by Sex	107
TABLE 3	Reported Effective Incidents by Class Standing	108
TABLE 4	Reported Ineffective Incidents by Sex	108
TABLE 5	Reported Ineffective Incidents by Class Standing	109
TABLE 6	Behaviors - Personal Relationships With Students	111
TABLE 7	Behaviors - Classroom Administration	112
TABLE 8	Behaviors - Classroom Presence	113
TABLE 9	Behaviors - Organization and Presentation of Material	114
TABLE 10	Behaviors - Evaluation of Student Performance	116
TABLE 11	Evaluative Behavioral Statements	119

APPENDICES

Page

APPENDIX 1 - Tabulation of Effective and Ineffective Incidents by Category and Sub-Category 134

APPENDIX 2 - Number of Incidents Reported by Major 140

APPENDIX 3 - Categories of Incidents Reported by Major 141

APPENDIX 4 - Behavioral Evaluation Statements Derived From Other Critical Incident Studies 149

DEVELOPMENT OF AN INSTRUMENT TO EVALUATE COLLEGE CLASSROOM TEACHING EFFECTIVENESS

Introduction

The general area of personnel performance assessment is one of the most difficult in all of Psychology, in particular, where there are few or no tangible products resulting from the performance effort. This condition describes the area of college teaching. In the classroom, a course is "taught", student learning usually is evaluated with some sort of test but there is little to indicate the contribution of the teacher to any learning that may have taken place. It is possible that learning occurred in spite of, or because of, the particular teacher.

Despite the difficulty of measuring classroom teaching effectiveness, there has been the mentioned tide of criticism concerning the educational system and, at least implicitly, the quality of teaching. More specifically, there has been established the "Project To Improve College Teaching", by the American Association of University Professors and the Association of American Colleges with the support of the Carnegie Corporation. The project is concerned with recruitment of qualified persons to teaching, assessing their development, and maintaining their effectiveness as teachers (see Eble, undated). In general, the major interest seems to be mainly concerned with the effectiveness of college teachers.

As a basis for any "improvement" of college teaching, recruiting or training qualified personnel, or any other such personnel actions, it is essential that a performance assessment method be developed as a requisite for determining who is and who is not an effective teacher. Actually it is not likely there will ever evolve a dichotomous evaluation as implied in the previous sentence but that persons will vary by degree of effectiveness as teachers. Further, in such complex work, there are likely to be several performance dimensions and probably no specific individual will be outstandingly effective or ineffective in all of these. In fact, it is possible that some dimensions can be mutually exclusive in that if a person is effective on one dimension he is necessarily ineffective on another. In any case it is necessary to develop an assessment procedure for a given performance before it is possible to recruit and train for it or devise methods to improve it. As has been shown, Ronan and Prien (1966, 1971), no systematic personnel procedures are possible without there being available some adequate criterion to evaluate the performance in question.

It is the intent of this research to detail needed first steps in the development of some conceptual criteria and procedures to measure effectiveness on the several dimensions of teaching performance by college faculty.

SOME RELEVANT RESEARCH

This literature review is, by intent, limited in coverage. Any comprehensive review of the literature concerned with the evaluation of teaching would require years of effort as shown by the bibliographies compiled by Barr & Jones (1958), Dorcas & Tiedman (1950), and Eells (1967). The latter comprises some 2,300 titles. In addition, the major effort of research on teaching effectiveness has been concerned with primary and secondary schools with results which may or may not be applicable to college teaching.

This concentration of research effort has been shown in a comprehensive literature review by Morsh and Wilder (1954). These authors reviewed some 900 primary sources and included in their final review almost 400 as selected by their criteria for presentation.

The monograph discusses criteria for teaching effectiveness and the relation of various pupil and teacher characteristics to these criteria. Some of the more important results from the studies might be summarized as:

(1) The wide variety of measures that have been employed in the various studies and, in particular, the lack of replication of most of the findings.

(2) Ratings of teacher effectiveness tended to be reliable but were not related in any substantial way to objective measures of teacher performance. In particular, ratings by administrators show low correlations with objective measures, for example, student "gains" as measured by various tests.

(3) The difficulties of using student gains as criteria were pointed out, statistical problems receiving the most emphasis.

(4) Predictors of teacher effectiveness such as intelligence, college grades, various "national teacher tests," aptitudes (Knight, Coxe-Orleans, Stanford), and personality measures showed varied and tenuous relationships with any criteria.

(5) A suitable criterion for teaching effectiveness must take into account student gains, this is the objective of teaching, the measure should be objective (here the possible utility of controlled observations is stressed), and a composite or global criterion of teaching effectiveness is, as of now, unlikely.

(6) Prediction of teaching success and teacher training will only make progress as a suitable criterion of teaching effectiveness is developed.

In reading the report, one is struck with the tremendous amount of effort that has been expended on teaching research and, at the same time, the lack of real progress in the area when the report was presented and continuing to the present as shown by the research subsequently presented.

Ryans (1960) in a research program concerned with primary and secondary teachers found three general areas around which more effective teaching seemed to center. They were:

(1) Warm understanding, friendly vs. aloof, egocentric, restricted behavior.

(2) Responsible, business-like, systematic vs. evading, unplanned, slip-shod.

(3) Stimulating, imaginative vs. dull, routine classroom behavior.

As will be seen, these areas bear considerable resemblance to some found for college teaching; however, the research program also resulted in findings that make generalizations to college teacher populations somewhat questionable.

Some examples of such findings -- women teachers were rated higher than men teachers, particularly at the secondary level, teacher age was related to evaluations, teachers in larger schools were rated higher, relations of teacher behavior to pupil achievement were close in primary schools, only the "stimulating - imaginative" teacher rating affected pupil achievement in the secondary schools, and particular courses indicated students achieved at a higher level (mathematics and science).

Other interesting findings of Ryan's research were that the more highly regarded teachers had had some form of teaching experiences as young persons as, "teacher asking them to take a class" or "reading to children." The better teachers showed higher emotional stability and those rated high on all scales were rated high by all raters, including administrators, and were more intelligent and stable. This latter point has been found in many rating studies, that is, the really outstanding performer is recognized by all. In addition, the ineffective teacher tended to be self-centered, anxious, and showed "restricted" behavior in the classroom.

These latter points may apply to college faculty but there is as yet insufficient research evidence to indicate such is the case. In general, the above is cited to indicate why the research on teaching largely is restricted here to studies of college teachers. In any activity as complex as college teaching, there are very likely a great number of possible performance dimensions. As an analogy, studies of physicians, Taylor et. al. (1964) have shown between 25 and 30 factors on data describing performance not including diagnosis and actual treatment. The inference is that professional work might include as many as 50 possible job performance dimensions. One area of the college faculty member's job performance is that of classroom effectiveness, which may comprise several specific performance dimensions, and which has received some research attention. The material presented in this section is intended to describe some of the research literature and the attendant problems concerned with assessment and evaluation of the effective teacher.

As in all organizations, the performance of job incumbents in faculty positions is constantly being evaluated on a more-or-less formal basis. On the basis of such evaluations, administrative actions such as promotions, salary increases, and others are made.

A study by Gustad (1961) presented results of a survey by the Committee on College Teaching of the American Council on Education. The data were taken from replies to a questionnaire concerned with procedures and practices involved in faculty evaluation. All collegiate members of the Council were sent questionnaires and 50% replied. The replies were tabulated and cross-tabulated and the general findings were:

(1) The effective decision maker with regard to faculty personnel actions is the department head. In general, his recommendation passes to a dean, then to the president, and is approved. Critical in this procedure is the department head-dean meeting; all other actions were seen as "primary group" functioning, that is, the administrative officers are cohesive and support each other.

(2) Classroom teaching was said to be the most important factor in any evaluation. "Personal attributes" such as cooperation, loyalty, church membership and activity were of secondary importance.

(3) It was found that with few exceptions all evaluations were based on hearsay. The data sources were informal student opinions, formal student opinions (ratings), classroom visitations,

colleague's opinions, and opinions of chairmen and deans. It was pointed out that the validity of all these opinions is unknown. A quote from the study summarizes the situation: "It is apparent that little is done to obtain anything that even approaches sound data on the basis of which reasonably good evaluations of teaching can be made. This being so, complaints that classroom teaching is paid only lip service must be to a considerable extent accepted as correct." (p. 205)

(4) Evaluation of research production was found wanting in that, "research is not, in fact, actually evaluated; it is counted." (Gustad's emphasis)

(5) Public service by faculty members was found to be largely evaluated by self-report of faculty members to department heads or deans.

(6) Finally, "extra diligence" by the teaching staff was evaluated by student counseling, advising, committee work, etc.

In general, the institutions reported they were dissatisfied with their evaluation methods or satisfied, about one-half in each category. The most often made suggestion by the respondents was that what was needed was some method for evaluating classroom teaching.

The entire study indicates that the performance evaluation of college faculty leaves something to be desired. A quote from the study makes the point quite adequately:

> "In general, to call what is typically collected or adduced to support evaluative decisions 'evidence' is to stretch the meaning of that honored word beyond reason." (p. 208)

Gustad (1964) later completed another similar study and found that the situation had changed very little. The most significant result of the second study was that the use of systematic programs of student ratings to evaluate faculty performance had declined. The use of committees to evaluate performance had increased leading to the posed question regarding just where such committees obtain information considered appropriate to assess faculty performance. The article goes on to point out that teacher performance evaluations are essential to sound administrative decisions, faculty self-improvement, and as research criteria but, if anything, such systematic evaluation is regressing.

Some suggestions (alternative evaluations) are offered to correct the situation. Student opinions are seen as valuable if they are asked the "right questions". For example, student opinion as to a faculty member's competence in his field is of doubtful utility but his answer to a question regarding his interest in a given course is relevant. Some other possible measures are delayed measures of course material retention, student performance in advanced courses, scores on standardized tests, in particular, "discrepancy scores", that is, actual as opposed to predicted scores in a course, and several other possible measures.

With regard to asking students the "right questions", a study by Langlen (1966) is pertinent. The questionnaire used in the study centered around questions concerned with "your interest in" and "your understanding of" the particular subject matter. The questionnaire is used annually at the University of Washington and some 500 "sections" and 15,000 students are involved. Participation in the program by the faculty is voluntary and the results are given to the administration only with individual faculty member permission. The results of assessments have yielded correlations ranging from 0.70 to 0.84 with student understanding and help from the instructor and, 0.56 to 0.86 between item scores and interest in the course. Obviously, this is not validation of the instrument but it is an indication that students can agree as to what their perceptions are as to course value to them in relation to teacher effects. Possibly "right questions" center around these student reactions to teachers.

Actually the teacher evaluation practices described are little different than that of industrial, military, or other organizations. Performance evaluation in almost any situation is either deficient or nonexistent in any meaningful way. The tragedy is, however, that performance evaluation does occur and is on an unsystematic basis that is inequitable to all concerned. The point cannot be overemphasized, evaluation does occur, the problem is to put the evaluation on some basis that is as objective and fair as possible and that involves the same performance standards for all those persons being evaluated.

There have been various formulations of the requirements needed to correct the situation as described above by Gustad. One of these, Cureton (1951), has delineated some important but often unrealized considerations involved in assessing teaching effectiveness.

In the context of a suitable criterion for assessing such effectiveness, Cureton pointed out that in the broadest sense it is impossible to establish a completely relevant criterion in the present situation because the ultimate goal of teaching is to prepare persons for their total future adjustment in life. Behaviors to assess this adjustment and presumed teaching effectiveness, occur long after actual teaching has taken place and thus are never related to prior education in any research sense. This situation requires deriving more immediate criteria of teaching performance if teaching performance is to be assessed. As stated by Cureton, the first need is a statement of clear and measurable goals for educational efforts. When such goals have been established, measured teaching effectiveness would consist of determining student progress toward these goals and, by inference, teacher effectiveness. The statement further indicates a need to establish what goals can legitimately be required of formal educational efforts, that varied levels of student evaluation will need to be tested, and there is the need to construct adequate measures of student progress.

Eckert (1950) discussed the general requirements of the teaching job to show that actual classroom activities are only one phase of teaching. The work also encompasses administrative duties, student counseling, community services, and research. In actual fact the teacher is also a staff member but it is generally agreed in virtually all discussions of this area that classroom teaching is the important job duty. Some possible performance criteria are presented by Eckert; these closely parallel a list presented by Gray (1969), and possible methods through which teacher behavior might be assessed. One suggestion, related to Cureton's point of ultimate criteria, was the suggestion of life-time follow-ups of students. However, the point is again made that teacher effectiveness can be assessed by the effects of teaching on student behaviors.

McKeachie (1963) in a review of the literature concerned with the procedures and techniques of teaching has pointed out that research studies in the field have, for the most part, yielded inconclusive or ambiguous results. The ultimate nature of the criterion of student change in the direction of desirable educational goals was stressed and the need for measures of just what is causal in bringing about such changes.

Mitzel (1960) in a discussion of the criterion problem as related to teaching effectiveness mentioned the accepted standards required of all criteria, e.g. relevance, reliability, freedom from bias and contamination, and practicality. He classified possible criteria into:

(1) Product criteria-largely educational goals and student progress toward such goals. The difficulty of making these operational, the need to confine them to immediate and practical measures rather than the broader social context were recognized along with decisions as to relative importance and, finally, the primacy of teaching as the most salient teacher task.

(2) Process criteria-relating to student-teacher interactions, "climate", rapport, individual attention, in general, how a teacher can relate to students.

(3) Presage criteria-teacher characteristics that are often assumed to be related to teaching excellence such as personality traits, intelligence, training and administrative evaluations.

Finally, Mitzel took issue with the assumption implicit in many studies that teaching is a unitary trait and he pointed out that the research evidence indicates a teacher-student-context interaction that would make for varied success depending upon how these sets of variables might differ in given circumstances. Generally, there was recognition of the complexity of the evaluation problem and, in particular, that the concept of teacher effectiveness has no meaning as a concept apart from some adequately documented performance criterion.

The most detailed and wide-ranging discussion of teacher effectiveness has been presented by Barr, et. al. (1953). As has been mentioned earlier, all educational personnel actions require defining a "good teacher" and these authors brought out a point most relevant to the present study -- what is required is not who is an effective teacher but determining what is effective teaching behavior. The authors cited Bloom's (1956) Taxonomy of Educational Objectives as a definitive statement of educational goals with respect to guiding research efforts; however, there was also recognition of the "staff" functions previously mentioned. It was recognized that statements as to effective and ineffective teaching are required and the main body of the article was a rather detailed description of the requirements for adequate research to solve the problems posed. The general formulation was to state some specific and measurable objectives and then assess teacher behaviors as directed toward the attainment of such objectives in all teacher task requirements. The three broad areas of teacher responsibilities were seen as work with students, tasks as a teaching staff member, and functioning with relation to the community at large. The authors were at particular pains to stress the complexity of the needed research, the basic importance of adequate measuring devices, and the requirements to design and carry out the needed experimental studies.

All of these studies were in unanimous agreement as to the complexity of the teaching job, its many and varied dimensions. There was also general agreement as to the most important job duty; teaching students so that they make visible progress toward accepted educational goals was emphasized by all. Thus, the primary objective of research should center around determining what teaching behaviors are effective in efforts to have students reach the desired educational goals.

This objective can be approached by several different methods: (1) determination of the level of student learning which would entail standardized conditions and measures of that learning, (2) an estimate of the degree to which the teacher establishes conditions essential to learning, a task fraught with all sorts of difficulties, (3) a determination as to whether or not the teacher is concerned with teaching as the adoption of new content, information, or methods, at best making it inferential that good teaching results, and (4) soliciting the opinions and observations of the teacher's students. This last seems to be the most promising since the students are in direct relation to a given teacher, undoubtedly do have opinions as to performance effectiveness, and probably can report these. The question then arises as to a feasible procedure for eliciting the required information from students keeping in mind the requirements of performance measurement, e.g. comprehensiveness, fairness, and objectivity. The intent of most reported research was the development or evaluation of such measuring instruments. In order that the research may be fully appreciated, some previous research studies have been explored.

Ratings of Teaching Performance

By far the most common procedure used to evaluate teaching performance has been the use of some sort of rating scale. In such studies the rating instrument employed has been the "Purdue Rating Scale for Instruction", the "Miami University Instructor Rating Sheet", or some locally developed measure, recently, some so-called "forced choice" scales have been developed.

An early study by Bousfield (1940) summarized the general problems and pertinent research related to college teaching as they were seen at that time. As will be developed, neither student opinions nor the problems seem to have changed over the course of the decades.

As summarized by Bousfield, students complained that their professors were deficient in personality, lacked sympathy for and devotion to youth and learning, and "annoyed" students by rambling in lectures and "riding" students. At the same time, professors were conducting research into their own teaching effectiveness and attempting to make both themselves and their institutions better instruments for learning. As an incidental the study pointed out, "The research man and the scholar tend to be poor teachers." This piece of folklore is still current and not only with students. The most important observation pertinent here was, "a primary difficulty in any attempt generally to promote effective teaching is that there are at present no satisfactory measures of the quality of college instruction and, in spite of this, more or less definite opinions do exist, and information on these opinions should be helpful providing it is used judiciously." (p. 253)

To collect such opinions, Bousfield had 61 undergraduate students list what they considered the five most desirable traits of a college professor. From these listings 16 traits were derived and 3 others concerning leading discussion, research, and scholarship were added to comprise 19 traits in all. These were then evaluated for importance on a 0 - 10 scale by 336 male and 71 female students at two different institutions.

Fourteen of the traits received ratings of 7.0 or higher with research (3.9) and scholarly reputation (3.22) being given quite low ratings. There was no real difference in the ratings by men and women, with "fairness" (9.19) and "mastery of subject" (9.00) being given the highest importance ratings. It is interesting to speculate whether or not a reputation for research competence and scholarly attainment would result in a lower rating of teachers by students. This is, however, just speculation at present but as Bousfield noted, "implies the double obligation of effective teaching and scholarly production."

As will be seen, this early study uncovered desirable teacher characteristics that are important to students today but does not indicate either the actual behavior or behavioral correlates for such a trait as "fairness."

An even earlier study than that above, Hulman and Armentrout (1936) used the Purdue Rating Scale to obtain ratings of 46 different teachers by 50 classes, a total of 2,115 ratings was obtained. Data were also collected on student personality (Bernreuter Personality

Inventory), ratings of instructors at Purdue, class size, severity of grading, sex of teacher, maturity of the student rater, and previous ratings of some teachers.

Wide differences among students rating the same teacher were found and teachers were also rated quite differently by different classes on individual traits and as a group. Standard deviations varied from 4.25 to 18.45 for the classes and from 4.75 to 27.30 for the individuals, on a 100 point scale. The standard errors of the means were found to be so large as to make the data relatively useless for evaluation purposes. Teacher differences in ratings were not meaningful because of these large rating errors, little difference was found in ratings of teachers from two different institutions, for experienced versus inexperienced teachers, or on the basis of age. None of the demographic or descriptive data seemed related to the student ratings.

An intercorrelation matrix for the 10 rated traits showed quite high intercorrelations with the exception of "Personal Appearance" which showed relatively low correlations with the other traits. Here is an example of the well-known "halo effect" in ratings, i. e., to rate a given individual near the same point of every scale in the rating form.

The same halo effect was quite obvious in a factor analytic study by Maslow and Zimmerman (1956), virtually all of the major loadings were on one factor. Remmers (1934) has also shown the existence of halo in student ratings and, peculiarly, college students tended to show the effect to a greater extent than did high school students.

In summary, the above studies have indicated that student ratings of teaching effectiveness are different for the same teacher by different raters, yield large standard errors which, in some instances, cover half the rating scale, and are further distorted by the halo error. In general, to determine any "true score" appears to be impossible with ratings.

To illustrate one source of error in such ratings, the data in Table 1 are presented. These results are from a study conducted by a student group at Georgia Tech who asked professors to allow their classes to rate them. The ratings were on a five point scale: (1) almost never occured to (5) almost always occured.

Table 1

Question Number	Distribution of Student Ratings for Selected Questions				
	1	2	3	4	5
16	1	1	0	3	20
17	6	5	4	6	4

Question 16 concerned the person knowing "more about his subject than just what was in the book." To the author's gratification, 20 of 25 students rated this (5) or, almost always occured in his class. However, there are also ratings of 1 and 2, not so gratifying. It is, of course, tempting to believe 20 were right and 2 wrong, but is this so?

Question 17 asked about difficulty of copying "what the instructor put on the board." This is a virtually rectangular or chance distribution, and no judgment is possible.

Reported research concerned with ratings usually does not report item distributions but, if the above are at all representative, ratings would appear to be of quite doubtful utility.

A third early study approached teacher rating from a different point-of-view, "annoying habits" in the class room, Moore (1937). Using a list of traits developed in a previous study, 99 male and 133 female students at four different colleges, were asked to rate traits as "greatly" and "slightly" annoying. The six most mentioned greatly annoying traits were: (1) Rambling in lectures, (2) "Riding" students, (3) Pausing too long, (4) Using pet expressions, (5) Nervous movements, and (6) Writing illegibly on the board. The six most mentioned slightly annoying traits were: (1) Frowning, (2) Pausing too long, (3) Cocking head, (4) Sticking hands in pocket, (5) Twisting mouth, and (6) Walking about too much.

Students were in fair agreement from different schools as to the above and sex differences were minor. Generally, some of these "habits" are found in the research to be reported here but some are not. In particular, personal habits as the "slightly annoying" above were largely ignored. However, an interesting speculation would concern a study of "annoyances" as related to student ratings of professors. To the author's knowledge such a study has never been done.

As will be noted throughout this report, the emphasis is on student evaluation of teaching performance. There are two reasons for this. First, the students are in a direct relationship with teachers and can and do observe actual teaching and all its behavioral elements, whereas, other observers can only see a limited sample of a given teacher's behavior. Second, there is evidence to show that observers and students do not agree on teaching performance as in the following study.

The study is that of Cook and Leeds (1947) dealing with high school teachers. Teacher performance was evaluated using ratings by principals, "experts", and students. The rating intercorrelations were: (1) principals vs. experts, 0.48, (2) principals vs. pupils, 0.39, and (3) experts vs. pupils, 0.33. The results, obviously, indicate very little agreement among the various raters. This is a quite common finding in all studies of performance evaluation where similar methodology is applied. However, it will be shown in a later study, Douglas (1968), that student perceptions of teacher performance are related to student learning, which is the generally stated objective of teaching.

There have, obviously, been hundreds of teacher rating studies performed over the years, most of them were for personal or institutional use and have not been reported in the research literature. However, those that have been reported leave little doubt that ratings by students have, at best, a marginal utility. This conclusion is both explicitly and implicitly recognized in the research work that is cited below. The work is centered around determining student characteristics as related to ratings of faculty and attempts to improve the rating instrument and its interpretation.

Downie (1952) analyzed a 36 item rating scale, some 16,000 evaluation forms in total, covering over 400 staff members, with regard to various student and teacher characteristics. Some of the findings were:

(1) Students with lower grade point averages gave lower ratings on all items except one than did the students with higher averages, most noticeably on class testing procedures.

(2) Whether the course was required or elective had relatively slight influence on ratings.

(3) Upper classmen gave higher ratings than did lower classmen on items centering around new interests and intellectual curiosity.

(4) Class size seemed to be the most influential contingency variable. Large classes (over 30 students) gave lower ratings to instructional procedures, tests and quizzes, and value of the course. In all, 17 items of the 36 total were rated lower. In the smaller classes, instructor-student relations items received lower ratings. These included the background of the instructor, his knowledge of subject, impartiality, and use of sarcasm and ridicule.

(5) With regard to faculty characteristics, age did not affect ratings, rank was related to rating, and graduate degree holders were generally seen as more effective with material presentation and gave more appropriate assignments, those with PhD's knew the subject matter better and were more stimulating, and were generally judged as more adequate. Years on the teaching staff and sex of teacher were not related to ratings.

Here, as in the studies below, there appear to be various contextual and situational variables that have some effect on student ratings of faculty.

Anikeeff (1953) reported a study of teacher ratings as related to "grading leniency", class attendance and student class level. In the study, 19 faculty members were rated on 8 traits by at least 50 students and/or three classes, in total, some 1,500 ratings. It was found that grade leniency was significantly correlated with ratings at the freshman-sophomore level but not at the junior-senior. Class absences were negatively correlated with ratings for only the total group. These relationships, found in spite of the small N in the study, are the more impressive for that reason. The indication is that student impressions of faculty may be related to some specific behaviors not directly related to teaching effectiveness.

A series of studies initiated by Bendig (1952) used somewhat more sophisticated statistical techniques to analyze the relationships of student characteristics to faculty ratings. One of these studies, Bendig (1953a), was concerned with the relation of course achievement to ratings of six psychology teachers by 132 students. The article points out that there is considerable research evidence to show that perceptions of objective stimuli are distorted by personal needs and traits and, further, that students could react to low grades by "punishing" an instructor even though his grade may be due to his own lack of aptitude or motivation.

Ratings for the study were obtained by use of the Purdue scale to give ratings of the instructor and of the course. Grades, at the time of the rating could be determined since the students signed the rating form. Grades were based upon three departmentally constructed multiple-choice achievement tests of 50 questions each (reliability = 0.88 for the average grade.) There were then three variables, student achievement, instructor rating, and course rating.

The three variables were tested (chi-square) and both normality of distribution and homogenity of variance were shown. Simple analysis of variance was then employed to test for difference between classes. Such differences were shown for both instructor and course ratings, but achievement test scores were not significantly different between classes.

The next step was to control achievement level in order to test for differences of ratings using analysis of covariance. It was found that instructor and course ratings were still significantly different using this technique. The differences were due to non-rectilinear relationships between means of achievement and means of ratings. This difference was due mainly to one class where the ratings were much higher than would be indicated by the regression line of obtained ratings and tests. Finally, the correlations of achievement and ratings (with non-linearity of class means controlled) showed that achievement and course rating were significantly correlated (0.28) but achievement and instructor rating were not (0.14). It was concluded that student achievement does affect rating but only to a limited degree.

Another study by Bendig (1953b) was based upon a previous factor analytic study of the Purdue Scale. Three factors, "halo", "instructor competence" and instructor "empathy" were found, Bendig (1952).

For the study the procedure reported above was used to collect data but only sophomore scores were used to control for class level. Factor scores for competence and empathy were derived from the previous factor study and both analysis of variance and covariance were used to assess the achievement-rating relationships. The findings of the study were:

(1) Students can reliably discriminate instructor competence and empathy.

(2) Student achievement and rated competence are highly and negatively related but there is no relation to empathy.

(3) Mean ratings were not related to sex but there was a significent sex-instructor interaction for the particular trait of instructor competence. Some are rated higher by women, others by men, and this is independent of achievement differences between men and women.

(4) Three-way interactions of competence, instructor, and student sex were also significant.

A final study in this series, Russell and Bendig (1953), was concerned with the relation of aptitude, achievement, and teacher ratings. The study covered 231 students in introductory Psychology, six teachers, and used the quantitative and language scales of the American Council on Education Psychological Examination as predictors. The rating of instructors was accomplished using the Miami Instructor Rating Sheet. The students were divided into groups of those whose predicted course grade was above that predicted (plus), equal to predicted, and below that predicted (minus). Using scores, two achievement ratios were calculated for each instructor (1) plus and equal students divided by equal and minus students and (2) mean of algebraic differences between predicted and obtained grades for each student.

No overall difference in instructor ratings was found for the plus and minus groups but when relative achievement was held constant, differences between the groups on single scales of the rating form were found. Plus students rated the items describing the course more favorably, as text, examinations, overall rating, but there was no difference in attitudes towards instructors. The ratings were not significantly correlated with the two achievement ratios.

In general, the series of studies has shown that instructor ratings by students are related to course achievement and certain other student characteristics. The effects shown are somewhat limited but, then, the included variables are somewhat limited. In any case, the studies do open to question the uncritical acceptance of ratings as being adequate measures of teaching effectiveness.

Elliott (1950) derived a list of teaching criteria based upon work by Barr (1948) and set out to analyze them in two separate studies.

The first study collected data from several colleges and universities in Indiana and data were obtained for hundreds of teachers and from thousands of students.

The criteria used in the first study were the Purdue Rating Scale and How To Teach And Learn In College test (attitudes toward various teaching practices). Along with the criteria data, demographic information was collected from student raters, e.g., class, veteran-non veteran, sex and grade (upper or lower half) in the particular course.

The results of the study showed that graduate students tended to give higher ratings than undergraduates. No other student characteristic was related to the ratings given. Characteristics of teachers on the attitudinal measure showed no differences by rank or experience and only a slight relation to level of higher degree and sex of the teacher (women scored somewhat higher).

As to ratings received, sex of rater did not affect ratings given, those teachers with low experience received low ratings but high level of experience did not necessarily result in high ratings, holder of bachelor's degrees only were rated low but there was no difference for the advanced degrees. Three of the rating scales were said to differentiate between teachers in the "expected direction" but the actual data were not presented.

The second study involved only instructors in the Chemistry Department of Purdue University. For the study two items were added to the Purdue Rating Scale, one compared the particular instructor with others the students had encountered at Purdue, and the other asked if the occasion arose should the particular instructor be replaced. The attitudinal test was given to the teachers involved plus an elaborate subject matter test in Chemistry. As another criterion, each student was given a "discrepancy score." This score was derived by comparing actual grade received in the course versus a grade predicted from a multiple regression equation based upon the American Council on Education Psychological Examination (students were graded on several tests in both laboratory and recitation work that were primarily objective in their scoring.)

The results of the study indicated only five of 24 possible correlations between ratings and discrepancy scores were significant and these were rather low. Neither subject matter knowledge of the instructor nor teacher scores on the attitude scale were related to student achievement. More detailed analyses indicated that some teachers were possibly more effective with certain students but, in general, the results of both studies were negative both with regard to the correlates of student ratings and student achievement.

A somewhat later study by Weaver (1960) assessed the effect of student's expected grades and ratings assigned to instructors. In the study, 12 instructors in a variety of subjects were rated by 39 classes, a total of 699 students. The rating form consisted of seven scales related to instructor personality and 19 scales of teaching ability. Each was evaluated with a five-point scale. Every student was asked to indicate the grade he expected to receive in the course.

The results indicated that instructor ratings were biased in the direction of the grades they expected to receive (t tests). Generally the bias was related to teaching ability rather than the personality variables, the latter were not demonstrably different by expected grade. Students expecting to receive lower grades seemed less discriminating in their ratings than those expecting higher grades (smaller rating standard deviations).

A study that was somewhat more complex than most of those reported was that of Rayder (1968). The study was planned in view of recent campus developments regarding student desires to see the quality of college teaching improved and points out that there is "only one ultimate source of its improvement - the individual teacher himself." (p. 77) The need for research in the area is stressed.

The actual study was conducted using a one page rating form based upon findings by Ryans (1960) _op. cit_. The scales were constructed using dichotomized items such as aloof vs. friendly, slipshod vs. systematic, and routine vs. stimulating behaviors. Each was evaluated on a 7 point scale. Both student and faculty demographic and behavioral data were obtained for some hundreds of students (faculty N not reported).

It was found that student's sex, age, grade level, major area, and grades previously received (GPA) from an instructor were _not_ related to ratings nor were multiple correlations of these variables significant with the rating criterion. Multiple correlations of student ratings with instructor variables were higher but the highest was only 0.27, practically of little use. There was some suggestion that older, more educated and experienced teachers received _lower_ ratings but again the relationships were quite low. In general, student ratings of instructors are stated not to be affected by student characteristics; however, partial correlations were not presented and might have shown some effects as in the Bendig studies above.

A unique study by Webb (1967) was concerned with student and faculty ratings of 13 course goals, such as teaching of knowledge, development of intellectual skills and abilities, etc. The major purpose was to estimate student and faculty correspondence in rating these goals for "emphasis" given each goal, "performance" or accomplishing the goal, and "relevance" or importance of the goal. Fourteen instructor's classes totaling 324 students participated in the study. Faculty rated themselves on each goal and students rated the teachers. All ratings were anonymous.

In terms of correlations, comparison of student ratings and instructor self-ratings for emphasis and performance showed only 3 of 28 possible correlations statistically significant (Table 4 in the report). By classes there was considerably greater agreement, in particular, for emphasis and relevance where 18 of 24 correlations were significant. With performance, only 4 of 12 were significant. Correlations of student grade point averages with ratings of instructors were negligible.

The general implication of the study was the considerable lack of congruence between what instructors say they are attempting to do and student perceptions of these attempts. Of particular interest here is that the findings with performance ratings leave much to be desired as evaluative devices. The need is for both students and teachers to come to some better understanding as to teaching activities and their intended goals. In broader terms, the implication is that student ratings can be based upon inadequate perceptions of both teachers and course goals and therefore such ratings would lack relevance as criteria of teaching performance.

An implicit recognition that student rating of teaching performance was not truly acceptable as a criterion has been shown in the development, parallel with rating studies, of a different rating form - the forced-choice. The potential usefulness of the technique for teacher evaluation was suggested by Ryans (1954).

Ryans pointed out that teachers at all levels have resisted ratings of their performance. The opposition has been based largely upon the qualifications of the persons who will judge their performance and the unreliability of ratings. At the same time Ryans indicates that judgments do occur and that they are absolutely essential to any personnel program if selection, placement, and training are to be done with any justifiable basis. Essentially, if the characteristics of good and poor performances are not known, how can any personnel action be justified?

The article goes on to suggest that the forced-choice sort of rating instrument is designed to have observers choose between behaviors describing the person being evaluated. The behaviors are often made equally social desirable (scaled) and are arranged in groups of two, three, or sometimes more as:

This teacher is:

Fair	
Democratic	
Responsive	

Respondents might be faced with an item as that above and asked to mark one (M), most characteristic, and (L), least characteristic of the particular teacher. Various formats and combinations of behaviors are obviously possible with the technique.

Ryans then describes the Classroom Observation Scale developed under his supervision constructed as the above. He mentions that the behaviors were obtained for inclusion in the scale through the use of "critical incidents", a technique used in the present research. It is pointed out that the developed scale avoids many of the problems of the graphic ratings usually employed and appears promising as an evaluative device. As presented by Ryans, the instrument was intended for use by raters, as peers or supervisors, not by students. The need for trained raters and a systematic assessment program were discussed.

From the time of Ryan's presentation to the present, research studies have continued using the forced-choice technique. An early study was that of Lovell and Haner (1955).

This research was one of the earliest to recognize the need to obtain student opinions describing desirable and undesirable teacher behaviors. Senior students (N-248) were asked to write essays describing the best and poorest college teachers they had encountered. From 200 essays elicited, a list of 53 behaviors describing the best teacher and 54 describing the poorest was culled. Seniors were then asked to check these items as descriptive of best, average, or worst teacher. From these returns, the items were analyzed for a preference index (applicable as descriptive) and a discrimination index (degree of item differentiation through biserial correlation.) The items were grouped into tetrads, "tried out", and a final list of 18 tetrads retained. The scale was then administered by the entire faculty with a procedure that allowed faculty members to remain anonymous.

For the scale, test-retest and odd-even reliabilities were found to be close to 0.90. Administered several years later, the preference and discrimination indices showed almost perfect correlations for the two administrations. As an indication of validity, the method of construction was thought to ensure this characteristic. Faculty members also completed the scale thinking of "flesh and blood" people, not their own teachers, and their results showed high correlations with students on preference (0.74) and discrimination (0.92), with cross-validation. All this was taken as indicative of scale validity.

In addition, studies were made asking students to deliberately "fake" the scale to assign high or low ratings. They were able to do so to some extent but the results were statistically insignificant. Studies by class size showed that classes of 31 or larger gave lower teacher ratings, that men and women gave similar ratings, ratings in required classes were lower, and seniors gave higher ratings than other classes. No general conclusions were stated in the research report but it appears that a carefully constructed scale based upon student opinions is technically and operationally more sound than the usual graphic or forced-choice rating scale in the assessment of teacher performance.

Crawford and Bradshaw (1968) have presented a study somewhat similar to that described above. Some 300 students wrote themes describing effective classroom teaching and, from these themes, 13 statements describing such teacher behavior were isolated. For example, the highest ranked was, "Has thorough knowledge of subject matter plus substantial knowledge in related fields."

These items were randomly paired (78 pairs) and 158 students in psychology, 50 faculty members, and 30 administrators (61 of 80 total returned) were asked to choose the one of the pairs considered more essential or critical to the effective teacher. The items were scaled using the "modal discriminal process" for means and ranges by 10 different groups of judges as faculty by rank, departmental chairmen, male and female students, etc. It was found, using "coefficients of agreement", that there was substantial agreement for all judges concerning the items. Greatest consistency of opinion was found for Deans and Administrators, and lowest agreement among groups of female students. Further, chi-square tests of independence for groups, as compared to the total group of judges, indicated that the judgments of effective teacher characteristics differed among the groups. Intercorrelations of group judgments were quite high, indicating similar order of rankings, but scale values were markedly different.

Generally, groups were quite consistent in their judgments but the different groups of judges see different characteristics of importance in good teaching. It would seem that judgments of effective teaching need to be evaluated in light of the persons doing the judging and, further, no teacher is likely to appear effective to all judging groups.

Recently two other studies employing the forced-choice technique have appeared, Sharon and Bartlett (1969) and Sharon (1970).

The latter discusses some of the difficulties associated with ratings, in particular, leniency as is often shown (skewed responses) in performance ratings. The authors, in the former, set out to compare a forced-choice rating form with a graphic form. The forced-choice form was constructed from description of teaching effectiveness, with preference and discrimination indices found for the items, the Sharon (1970) study. The rating scale was a form in use in the institution. Both scales showed acceptable reliabilities.

Ratings were made under four conditions with each instrument, (1) where the results were to be used only for experimental purposes, (2) results to be used by a superior for evaluative purposes, (3) the rater was identified, and (4) the rating had to be justified by the rater to the ratee. A control group was also included in the study who rated without instructions or directions as to the use to be made of the forms.

Under the conditions stated, the mean ratings were not significantly different using the forced-choice form but the graphic were. With the latter the means under conditions (2) and (4), above, were more favorable but did not differ from each other. All of the graphic ratings showed significant negative skew. For the groups the correlations of forced-choice vs. graphic were, control vs. control (0.58), the two evaluative groups (0.80), the two identified groups (0.85), and the two justifying groups (0.46). Another point was that some raters resisted responding, that is, omitted some of the forced-choice items.

In general, the forced-choice instrument seems more useful, in particular, where results of ratings are to be used for evaluative purposes or where raters are identified. This, however, raises the question of the sincerity of raters and the potential for mischief through deliberately distorted ratings. If raters do change their ratings where they are known or where the results are to be used in an administrative way, is it possible ever to obtain useful ratings ? If evaluations are to be used "experimentally" forever, no progress will be made and, further, which rating is the "true" score ? It would seem the need is to study the behavior in question rather than methods of its assessment.

Taft (1959) has recognized most of the problems inherent in a faculty evaluation system as:

(1) Evaluative criteria are not established (agreed upon ?) by all the persons involved.

(2) Some criteria are not explicitly stated.

(3) Relative importance of the criteria to institution objectives is not made explicit.

(4) The relation of an evaluation to the institution objectives is not stated.

(5) Evaluation is by subjective judgment of a limited number of evaluators.

(6) Standards of performance are not explicitly stated.

(7) Incentives are not provided for faculty to move toward institution objectives.

Taft's proposal is to have a committee established to provide these desirable attributes and, on a scale, rate each one. Through use of decision and utility models, a given person's value to the institution could be established. As has been shown, administrators and faculty do not agree with students as to teaching excellence and furthermore, the rating is one step removed from the classroom where behaviors can be observed. In addition, this "expert opinion" method has been found wanting both in committees failing to agree on values or even what to evaluate. In fact, it has been shown in other areas that such managerial and administrative personnel do not know what their subordinates are actually doing, Prien (1962) and Prien and Powell (1961).

The use of ratings to evaluate teaching effectiveness has been questioned by Kent (1966). Some of his objections were that the forms have too many questions, lack organization, and usually contain some rather vague terms. Ratings by students are not objective in that grades, class level, and other contingent and contextual variables affect teacher ratings.

That such criticism has some justification received early support in a study reported by Starrak (1934). A three item graphic scale, in use for some years (40,000 ratings), had shown that teacher ratings gradually tend to become higher over the years and a correlation of 0.15 was found between student grades and instructor ratings. Ratings in the medium-sized classes (8 to 49) were higher, class standing has no effect on ratings, and a mean correlation of 0.47 between ratings on the three traits on the scale was an indication of the halo effect.

Class size, as related to teacher ratings, has been studied many times but Holland (1954) has reported an unusually well-controlled study bearing on this persistent and unsettled question. The research instrument was designed for student assessment of an instructor "image." Questions (10) were on a five-point scale and of the type asking, "considerate of others" or "self-centered." A total of 613 student respondents, fairly representative of the entire student body, constituted the research sample.

For the study, four instructors taught classes that were (1) large lecture-discussion (100 students), (2) medium sections (50 students), (3) small (less than 25 students), and a section in which each instructor "was allowed complete freedom to use whatever methods of instruction he desired."

The results indicated that variations in the evaluations were usually due to one instructor, that is, pooling sections for research purposes is a dubious practice. This is likely the same interaction effect reported by Bendig (1953) op. cit.

Other results were that the most favorable image came from the small lecture-discussion groups and there was a "strong tendency" for medium and large groups to produce the least favorable image. Instructors were not able to predict their image as reported by a class and, the author stated, interactions of variables were probably crucial to results and it is likely that various optimal class size - instructor relations will be found.

There are other objections to ratings that have been summarized by Ronan and Prien (1966) but probably the most serious is a lack of relationship between ratings of performance and objective indicators of the same performance.

In teaching, one study by Borg and Hamilton (1956) is directly relevant to this point. The study was conducted in an Air Force basic training school and involved 89 instructors. The instructors were rated by students, made self ratings, were rated by their supervisors, and by other instructors (peers) on varied traits by each of the several raters. Twelve problems, taught in the training, were used as criteria of performance as, escaping from a prison, and were solved by six man teams. Raters were rotated during the study, the ratings were anonymous and supervised, in fact, every precaution was taken to obtain acceptable data. The rsults showed the supervisory, peer, and self ratings agreed with an average correlation of 0.71. However, student ratings averaged only 0.17 correlation with the three other ratings. Further, the correlations with the performance tests were, students -- test (0.19), peers -- test (0.11), supervisors -- test (0.13), and self -- test (0.01). Generally, instructors and staff members showed good agreement on rating instructors (reputation?), students were much lower, and none of the ratings were significantly related to the performance measures.

Others have been a study by Rush (1953) showed that sales manager's ratings of salesmen were not related to actual volume of sales. Turner (1966) has shown that ratings of foreman performance were not related to turnover, grievances and other indicators of performance but tended to fall on one factor. Ronan (1963), with skilled tradesmen, found four factors among 11 indices of performance, one of which was a rating-promotion factor. Other indices were independent of ratings. The studies cited above were all factor analyses of matrices containing objective and subjective (ratings) measures of performance and the ratings and objective measures of the same performance were found to be orthogonal (independent) of each other in every case. This finding is quite general, although only a few of the studies have been cited, and the only conclusion is that reported ratings of performance are not related to actual objective performance measures. There is, in fact, one study by Lifson (1953) that demonstrated more variance in ratings was due to the raters involved than to the actual performance variance. In general, ratings of teaching performance have shown little or no relation to evaluating of student performance.

There are other objections to ratings for which there is little or no research evidence but that seem to present real difficulties in the use of ratings to evaluate and assess any performance behavior.

Probably the most serious of these is somewhat peculiar to teaching, a consideration of professor reputations. Every faculty member acquires a reputation among the student body whether deserved or not. As previously mentioned, and shown by Rosenthal (1966), people tend to see in situations what they expect to see. It is possible that distortion of ratings due to reputations is quite serious and more widespread than is realized. A possible indication of this might be inferred from a study by Lacognata (1964). The study employed a 53 item questionnaire covering a number of possible teacher behaviors with provision made to indicate the "desirability" of the behaviors on a five-point scale. Faculty extension teachers (75) and 96 full time evening students completed the form.

While there were some differences between faculty and students and between undergraduate and graduate students (t tests), there was general agreement as to the desirable "role" of faculty members. Actually, this very likely indicates some stereotype of the "typical professor" and deviations from this perception are likely to result in acquiring reputations of one sort or another. In addition, such stereotypes probably to some degree serve as a standard in ratings by students, although empirical evidence for this supposition is lacking.

Another serious objection to ratings is that the standard for the rating is often quite vague. If anything is to be evaluated there must be some agreed upon standard against which the evaluations will be made. In ratings, this usually boils down to each rater formulating his own standards because he is not given any explicit standard or the one he is given is so general as to be useless for any actual rating purpose.

Anderson and Hunka (1963) have discussed the general problem of teacher evaluation and pointed out that little in the way of definitive results has been forthcoming from the many research studies. They have emphasized the interaction of teachers-students-programs-situations as leading to an extremely complex evaluative problem situation.

The key point in such research domain must be to show the relation of the extent of student growth, along desirable intellectual and attitudinal lines, to teacher behaviors. With regard to the latter, it is pointed out that "pupils, evaluators, and administrators consider quite different attributes in conceptualizing the competent teacher."

These authors question quite seriously the ability of evaluators to rate other persons. Drawing from studies on human perception, it is shown that each person has his own idiosyncratic perceptual system, further, "perceptual defense" requires persons to organize, and thus simplify, their perceptual field and this results in a unique evaluation from each rater. In addition, the evaluator feels threatened and anxious by the need to evaluate and, as a result, can make quick, erroneous decisions in order to escape from the situation.

As Anderson and Hunka have seen it, the need is to study raters, ratees, and the interaction complexities in order to determine exactly what is being rated and what meaning, if any, such evaluations may have. The implication of the conceptualization of the rating system is then, that while ratings have had a wide usage as evaluation devices, there is no real knowledge of their exact value and will not be until systematic research is undertaken to determine the relevant situational parameters. This general point has been discussed by Ronan and Prien (1966) under, "Is Observation of Job Performance Reliable?", and their general conclusion was that ratings of performance are, for all practical purposes, useless. Further, they are likely to do an injustice to a ratee because they imply that an evaluation that is accurate and objective but it is, in fact, unreliable opinion. Finally, rating forms or scales are rarely based upon a wide sample of the behavior of interest; often such scales are "made up." This is a failure to recognize the complexity of the behavior involved, as well as an attempt to simplify the evaluation of what is an extremely complex phenomenon. Probably the main reason for forced-choice scales being seen as superior to graphic scales is that their authors usually select student opinion in their construction.

McKeachie (1968), in a general discussion of teaching and teaching methods has stated this complexity. What is being dealt with is a matrix composed of teacher-course material-teaching method-individual differences in ability, interest, and motivation along with many other situational and contextual variables all of which may affect both teacher and student behaviors. There is a lack of multivariate studies to assess *all* the variables and their interrelations, however, it has been repeatedly demonstrated that the better students learn better under any conditions which is what teaching is all about. However, as McKeachie points out, the teacher *does* make a difference in spite of teaching machines, programs, and all the other paraphenalia of modern technology. It would seem that if the problem of teacher evaluation is to be solved, it would be best to attack the problem from the viewpoint of what actual class behaviors are characteristic of good and poor teachers. The next group of studies has made some steps in this direction.

Factor Analytic Studies of Teaching Behavior

The studies presented in this section are largely based upon the factor analytic methodology and represent attempts to determine the "basic" dimensions of teaching behavior. They are presented in more or less chronological order to give an indication of knowledge development in this area.

An early factor analytic study was presented by Smalzried and Remmers (1943). The data were based upon ratings of teachers by 40 students using the *Purdue Rating Scales for Instructors*. This scale was originally constructed following a literature review by one of the authors (Remmers) and some colleagues. Ten traits composed the scale and it was generally agreed these were the important dimensions of teaching. The factor analysis followed Thurstone's centroid method and yielded two factors. Their actual independence is somewhat doubtful, for some items there were substantial loadings on both factors. The authors named the factors "Empathy Trait" and "Professional Maturity." The former implies "pupil-centered teaching" and the latter interest in and knowledge of subject matter plus a confident, stimulating manner of presentation. As pointed out above, the two factors do not appear to be completely independent but there is some indication of personal differences or reasons for being in teaching. On the one hand is a person trying to help students develop and on the other the intellectual or scholar who is teaching.

Crannel (1953) had some 300 students complete a course rating form containing 21 items evaluating the course being taken, some class administration concerns, and some teacher behaviors. Students were in several different subject classes and at all class levels. The procedure was to divide respondents into high and low groups, calculate tetrachoric correlations and, from these data, "cluster" the items.

Five clusters could be inferred but the author reduced these to three. They covered (1) "what the student thinks he is getting out of the course", (2) relationship students felt with the instructor, and (3) the effort the teacher appears to be putting into the course. Here, from 21 rated variables, were three basic teaching dimensions. It might be noted that, on a rating basis, there is considerable room for judgments to be distorted. For example, in cluster (1) the student might not be getting anything out of the course due to his own shortcomings but might, nevertheless, give the instructor a poor rating. Another possibility is that the highly competent instructor might not appear to put much into the course because of his competence. In general, the study appears to show some of dimensions for possibly student evaluations of instructors.

Coffman (1954) has reported a study of some 2,000 students rating 55 instructors on a 19 trait scale. Intercorrelations of the 19 variables were factor analyzed (centroid) and four factors extracted. The factors were:

(1) empathy-shown by ability to raise student interest, humor, student-instructor "feeling", tolerance, and liberality. This is not simple friendliness but an awareness of learning problems, patience, and an appreciation of the "world of the student."

(2) organization-reflecting presentation of course material (prepared) as well as interest and enthusiasm, punctuality, and self-confident ability to express thought.

(3) a "normal" individual who dresses neatly and is punctual and enthusiastic. Not related to overall teaching competence as are the other two factors.

(4) verbal fluency-ability to express thought, arouse interest, enunciation, and lack of personal peculiarities.

Intercorrelations of the factors were quite low, ranging from 0.17 to 0.10, indicating independence of the four dimensions described.

It is suggested that the study has provided data that could lead to a revision of the rating scale, a further analysis, and a less ambiguous evaluative device.

Gibb (1955) basing his analysis on work by Hemphill et. al. (1952), developed some 160 items to describe teacher behaviors under the nine categories of "leadership behavior" described by the Hemphill research. The form was completed by 119 male undergraduates to describe 70 instructors. Some described their first instructor in a given week, some the second and so on.

A centroid analysis (with oblique rotation of factors) yielded five factors of which four were interpreted.

I - ability to mix with students easily and sociably with lack of dominance and where a pleasant group atmosphere is encouraged.

II - "communication behavior". This seems to represent behavior of acting as protector or agent of the group and is not related to teaching competence. The factor is somewhat ambiguous but seems to represent a personal involvement in student learning and progress.

III - represent organized and clear class presentations along with innovative behavior in such presentations.

IV - a factor indicating behavior that encourages or forces students to learn. This is not necessarily the "iron fist" but the teacher does make it clear as to what he expects of students and is not likely to tolerate any "nonsense."

As an interesting summary point, Gibb pointed out the similarity of his results to those describing the better aircraft commanders. The suggestion was that there are effective and ineffective manners of behavior in social situations and relationships involving marked status differences. This important interpretation has been investigated by Lewis (1964).

Lewis had 169 students from three disparate disciplines complete the Guilford-Zimmerman Temperament Survey and a 100 item biographical information blank. They were then asked which professors had provided the "best learning experience" for them and to state the reasons for their preferences. Thirty-five faculty members in the respective departments completed the same two questionnaires. Analyses of both instruments gave largely negative results; student characteristics did not seem to be related to their selection of the more effective professors with regard to student learning.

Another related study, that of Gowin and Payne (1962), investigated the perceptions of teachers by students and perceptions of students by teachers.

A fifty item questionnaire containing desirable teacher behaviors was completed by 359 evening students and 14 teachers. The questionnaire format was:

		I think the teacher	I would guess the teacher thinks he
10.	Puts the student point of view in jeopardy as a way of making students think.	Yes	No

The headings of the Yes-No columns were changed for the teachers to, "I think I"and, "I would guess the student thinks the teacher--." In effect, a rating of teacher performance, a perception of the teacher intent, and a perception of presumed student opinions were obtained.

Students agreed on teacher level of performance but both groups misperceived the other, "--it appears that the students erroneously expected the teacher to overestimate his performance, and the teacher erroneously expected the student to underestimate his performance".

The teachers also completed an authoritarianism scale and the higher scores were negatively related to teacher expectations of the students seeing him as stimulating (r=0. 58). In addition, "The teachers who believed the students saw them as unstimulating were the teachers who assigned the lowest class marks (r= .83, p= .01) and who received the lowest average over-all ratings (r= .83, p=.01)." The teachers actually responded to perceptions they believed students held of them rather than to their self-perceptions, these latter were inaccurate according to the student evaluations.

There has been some research concerned with the social aspects of the classroom situation but the results here, along with those of Gibb (1955) above, seem to indicate that perceptions of persons in such situations can be seriously distorted by teacher characteristics and, further, there is the tendency for students and teachers to misperceive each other. If such is the case, the utility of teacher ratings can be even more seriously questioned than they have been in the past.

Cosgrove (1959) selected 200 items describing teacher behavior from an overall list of 900 discovered in a previous research study at Ohio State which had been conducted by Wherry (1950). These were sorted into categories by six qualified persons and the 150 items on which there was the most agreement were put into a rating check list. This check list was administered to 100 educational psychology students who were asked to think of a particular instructor and rate him on a 1-10 scale for each item. Preference and discrimination indices were computed for each item and the item intercorrelations were factor analyzed. Four factors resulted:

I - knowledge and organization of subject matter.

II - adequacy of relations with students in class.

III - adequacy of plans and procedures in class.

IV - enthusiasm in working with students.

The phrases were then grouped into 10 sets of four phrases each, with each phrase having a significant loading on a different one of the four extracted factors. Raters were asked to rank the phrases within each set for specific instructors (forced-choice).

The form was used by 12 sections of an educational psychology class, 8-12 students in each. The individual student data were averaged to result in two "profiles" for each section. Pooled "shape correlation coefficients" gave a value of 0.74 to indicate good agreement of the instructor ratings.

As the author is careful to point out, the developed instrument was not validated and comparisons between instructors are now warranted. However, a self-comparison can be justified on the basis of rankings of the various traits to indicate individual strengths and weaknesses.

Validation of evaluation instruments, as mentioned above, was attempted in a study reported by Solomon et. al (1964).

The study involved 24 teachers in 13 different schools teaching classes of evening students of 11 to 38 in size. Teaching behavior was measured using a 60 item student questionnaire, an 18 item scale asked teachers to report their motives and interests in teaching, tape recordings of two different class sessions were made, and observers rated teacher classroom behaviors using a 38 item instrument. To assess student learning, a multiple choice test was given at the beginning and end of the semester to measure factual and comprehensive knowledge of the course material (raw gain scores were used since they showed a high correlation with calculated residual gain). The items were intercorrelated and factor analyzed using principal components and varimax rotation methods. The factors extracted with percent of variance in parentheses were:

I -permissiveness vs. control (15)

II -lethargy vs. energy (11)

III -aggressiveness vs. protectiveness (10)

IV -obscurity, vagueness vs. clarity (8)

V -encouragement of content related (factual) student participation vs. non-encouragement of participation; emphasis on student growth (6)

VI -dryness vs. flamboyance (5)

VII -encouragement of students' expressive participation vs. lecturing (5)

VIII -warmth vs. coldness (5)

A basic finding of the study was that factual and comprehensive gain were related to only two of the factors, the former to factor 4 above and the latter to factor 6.

In addition to the objective learning criteria there were seven questions concerning student learning, interest in the course, and evaluation of the instructor. Of these seven, six showed high loadings on factor 4, two on factor 8, and none on factor 6. From this it can be inferred that clarity and expressiveness of presentation not only are basic to student learning but form the basis for instructor evaluation.

The authors also investigated some contextual variables such as class sex ratio, size, etc. but found only a few significant correlations.

An investigation was also made for possible non-linear relationships by trichotomizing the factor scores and relating these to factual and comprehension gains. Factor 1 was the only factor to show non-linear relations to the criteria. Teachers in the middle range of permissiveness vs. control had students showing higher gains.

The authors discuss the implication of their findings as, moderate permissiveness, energy, and flamboyance probably encourage students to become more involved in the course either on their own or by participation and factual gains are probably directly dependent upon "clarity of expression." This latter is also hypothesized as giving students feelings of intellectual security and confidence to increase learning. "Warmth" (factor 8) was related to two estimates by students of their own learning but not to the objective performance criteria; it is hypothesized that this is evaluation of the instructor as a person and has little bearing on learning per se.

A partial replication of the above study was completed by Solomon (1966) with 229 teachers in four institutions and a variety of disciplines. On the basis of the results of the prior study, the teacher evaluation here was accomplished by use of a questionnaire embodying 69 descriptive and three student evaluation items to be rated on a five point scale. The same factor analysis methodology was used and 10 factors extracted. They were:

I -lecturing vs. encouragement of student participation.

II -energy, facility of communication vs. lethargy, vagueness.

III -criticism, disapproval, hostility vs. tolerance.

IV -control, factual emphasis vs. permissiveness.

V -warmth, approval vs. coldness.

VI -obscurity, difficulty of presentation vs. clarity.

VII -dryness vs. flamboyance.

VIII -precision, organization vs. informality.

IX -nervousness vs. relaxation.

X -impersonality vs. personal expression.

Of the three items of student evaluation, only the over-all instructor rating showed relationships e.g., to factor I - encouragement of participation and II - energy and facility of communication. Again contextual variables were investigated with negative results; however, whether or not the course was "basic" or "applied" did show relationships. The latter showed teachers less nervous, critical, and clearer in their presentations. Differences in teacher behavior by intellectual discipline were also found. In general, the results here seem quite similar to those found previously in spite of different teachers, students, and methods of data collection. Particularly emphasized was that teachers in various disciplines probably behave differently because of value systems and/or the nature of the material to be taught. In any case, the studies seemingly have isolated some basic dimensions of teacher classroom behavior.

A final study in this area, Isaacson et. al. (1963), sought to relate teacher behaviors to personality variables. The teacher sample was 33 "teaching fellows" who were instructing for the first time and assessments were made in two separate years. These subjects completed a peer evaluation form of five personality characteristics, e. g., surgency, agreeableness, dependability, emotional stability, and culture by placing one-third of their fellow teachers at each pole for each trait. They also completed a 200 item self-descriptive adjective check list, and Forms A and B of the 16 PF Questionnaire (a commercial personality measure). Students of the subjects completed forms rating their teaching effectiveness (mean scores used in the study) but only the overall effectiveness item was eventually used in the reported results.

The only consistent relationship to both peer and student evaluations was "Culture", above. This finding was related to previous research, some of which has been cited in this review, i. e. Cosgrove, and possibly indicated some surface indication of general intellectual ability. However, a study by Ronan (1969) indicated no relationship between aptitude and personality factors and possibly what was found in the study here represents unique behaviors characteristic of some social stereotype of "the cultured person." In any case, this is one of the few attempts to determine the correlates of teaching behavior in terms of measured personal characteristics.

The studies presented above seem to be in good agreement as to at least three dimensions of teacher behavior. Several studies found organized presentation (clarity), warmth (empathy), and effort (enthusiasm) to be factors in student ratings of classroom performance. The question is, however, what behaviors represent these dimensions ? Specifically, what does one do to show his empathy for students ? The next section, presenting critical incident studies, attempts to isolate the specific behaviors in terms of effective and ineffective teaching.

Critical Incident Studies of Teaching Performance

The critical incident technique (CIT) was developed by Flanagan, (1954) for the specific purpose of creating performance measures to evaluate performance effectiveness. Basically, the technique gathers information from persons who observe performers in a given job. The observers are asked to describe incidents where they observed particularly effective and/or ineffective performance. Interviewing is continued until a usable incident is obtained. Usable is defined in terms of:

1. The behavior must be some objective behavior that all observers can agree did or did not occur. For example, a professor failed to meet his class three sessions in a row without informing the class members of his absence.

2. The behavior must be related to the aims and goals of the activity.

3. The behaviors to be collected must be gathered following identical rules and procedures for all interviewers.

Incidents are collected describing the behavior in question and then "categorized." The procedure consists of reading the incidents and sorting them into groups of similar behaviors. The categories are then named with regard to the behaviors described. With professors, for instance, there is likely to be a category regarding grading practices.

Two checks are available for the entire process. The first consists of having two or more persons categorize the incidents and then to calculate the percentage of agreement to give a measure of reliability. The second is to "hold out" a fixed percentage of the incidents until classification is completed. The held-out incidents are then read and placed into their proper categories. If the categorization has been adequate, no more categories should be needed for the new incidents.

This complete procedure results in a form that describes in objective and reliable terms, both effective and ineffective behaviors for the job or activity in question. Usually they are placed in a "Yes-No" format, that is, did the behavior occur or did it not? There is no inference or value judgment required of the observer. With the evaluative form developed it is possible to obtain an objective and fair evaluation of the job performance of any incumbent. The technique has been successfully used for many types of jobs, e.g., airline pilots, foremen, dentists, research scientists, aviation instructors and even ethical standards, see Flanagan (1954). The procedure will be used here to serve as a basis for construction of an evaluation device describing college teaching behavior. The device can serve as a criterion for research studies, show needed changes in performance (training), and give indications of the personal traits that could serve as a basis for the selection and development of college teachers.

Smit (1951) completed the first critical incident study concerned with teaching (Psychology). Incidents were collected from 497 students at two universities. Faculty members (25) observed lectures at one university and recorded incidents and, in addition, 11 of them supplied self-report incidents. All respondents were asked to record exactly what an instructor did that was outstandingly effective or ineffective in helping students learn, understand or apply facts, principles, or methods. Respondents also indicated why they thought the incident was effective or ineffective. A total of 2,342 critical behaviors was derived from 1,597 incidents and classified into 604 behavioral categories in three areas, (1) Presenting Material, (2) Estimation of Progress, and (3) Personal Adaptability. The material is much too extensive to present here but the research report points out that results create a basis for constructing instruments to evaluate teaching on an objective and behavioral basis.

Building from Smit's work, Konigsburg (1954) selected the 53 most mentioned incidents from the study and constructed a scale that could be answered in a yes-no fashion, (see Appendix 4). Another selection criterion for the items was whether or not the behaviors could be observed easily by students. The limitation of items was considered necessary in order to make use of the behaviors in planned experimental analyses. The total pool of items was considered much too lengthy in view of the limited class time available for evaluation. The list constructed could be completed by students in a range of 4-10 minutes and some ambiguities in the items were corrected after an experimental tryout of the check list that had been constructed.

Two forms of the check list (different directions to evaluators) were administered to four classes, one at the end of a three day period, the other daily for three days. Reliabilities were, respectively, 0.71 and 0.82. Certain items were revised from student comments (10 were eliminated) and the "three day" form was used because of time requirements. In the final test of the developed Instructor Check List, it was administered to eight sections of general psychology students, along with the Purdue scale previously described. The subject populations were five instructors and 243 students.

It was found that the correlations between the check list and scale were 0.35 for one administration and 0.20 for the other. Total scores on the scales for the same classes on the same day correlated 0.33 for the first administration and 0.25 for the second. In general, it is apparent that the two forms were not measuring the same thing. The study recommended further research work to determine the exact applicability of the two forms to teacher performance evaluation.

Another CIT study was by Douglas (1968) although the research is reported as in the area of "psycho-physics" and no mention is made of the work by Flanagan, Smit, or Konigsburg. This study went one step further than the latter's in that a validation procedure was incorporated into the research design.

The general aim of the research was to construct, administer, and evaluate a check list of teacher behaviors that were effective or ineffective with regard to student learning. Some 350 students reported incidents recording teacher behaviors that assisted learning and incidents that hindered their learning. Interviews with 80 other students gave similar incidents. A check list of 77 items, which were reported five or more times, was constructed. Actually three forms were constructed, all with the same behaviors but one had an overall rating of teaching effectiveness, one an overall rating of teacher ability related to learning, and one included both ratings. Items were further categorized into general effectiveness and learning items and for effectiveness (E) and ineffectiveness (I).

The three forms were administered, each to seven classes, with some 195 of each form. Scoring of the check list was by the ratio, number of effective incidents, number of ineffective, for each teacher. The learning criterion was measured by standardized final examinations administered to all classes, with student grade point average checked as a control.

The results indicated significant differences in effective behaviors were related to significant differences in all criterion measures, that is, achievement and overall ratings of effectiveness. Overall effectiveness rating as compared to the effective check list items was the best prediction. For the individual items, 49 of the 77 used showed discriminations between instructors at the .05 level or better. In general, such empirically developed behavioral scales seem more relevant to actual teaching behaviors than the more commonly used rating scales with regard to student learning.

Some of the problems alluded to in previous pages have been succintly summarized in a statement by Arden (1968). First, it was recognized that faculty evaluation by students, peers, and administrators does occur, and the need is to put such evaluations on a systematic basis. Secondly, it was recognized that students are the prime sources of relevant information concerning teacher performance. Finally, the point is brought out that specific and measurable information is a basic requirement in proper performance evaluation. In this connection, Arden presents a list of questions that in content and format duplicate some of those to be shown resulting from the research reported here. They are questions such as:

(a) Did professor X arrive to class on time and generally keep the class in session for the full period?

(b) Did he speak distinctly enough to be heard throughout the classroom?

(c) Did he return papers within a reasonable length of time?

The statements are objectively observable behaviors that can be evaluated on a Yes-No basis. As will be seen, this is the point of the present research.

In general, the problem of evaluating performance in some effective manner is endemic in all sorts of performance and is basic in the study of such performances. It is recognized as the "criterion problem" in Industrial Psychology and, more recently, has been formulated for Clinical Psychology, as shown in the recent statements described below.

Mischel (1968) discussed the applications to psychotherapy of studying behaviors, particularly his Chapter 8. It was pointed out that clinical assessments usually, "have ignored the individual's actual behavior in real-life situations." (p. 279) It was further mentioned that the usual behaviors sampled are from some form of psychometric device or interview whereas life behavior data are obtained by unsystematic methods. In particular, the latter are usually global statements and not the carefully defined operational descriptions that are needed. The goal is to design individual treatments related to specific behavioral problems that have been determined. To accomplish this requires detailed and specific descriptions of performance as it exists. The same point is made by Krasner (1971) where it is stressed that if behavior is to be assessed it must be known what behavior occurs and under what conditions. Here, as with teachers, the key is the study of performance behaviors.

If human behavior is to be more fully understood, it is a basic requisite that performance be studied and adequate measures of all performances developed. From such performance measurements it is then possible to infer or construct appropriate selection, placement, and training methodologies and devices. Without such performance measures, one is groping in the dark. The CIT seems to offer the basic methodology to collect the relevant data and to construct such performance measures. In the research reported below, the attempt was made with the CIT to determine all the dimensions of teacher behavior as seen by students not only those relating to learning alone, by collecting a wide sampling of incidents of "best" and "poorest" teacher performances in their general dealings with students.

RESEARCH PROCEDURE

In view of the doubtful utility of ratings as a method of performance evaluation and the indicated potential of the Critical Incident technique (CIT), it was decided to collect incidents from a wider sampling of students at the Georgia Institute of Technology than has been reported in previous studies.

The data were collected by students from the author's classes in Industrial Psychology during the Winter and Spring quarters of 1969-1970. Some 120 students collected 2663 incidents from their fellow students. The basic procedure was to use the following question format:

> "I would like you to think of the best (worst) professor you have had at Georgia Tech (do not name) and give me an incident that made you think this."

Two incidents were collected from each student interviewed, one describing the best and one the worst professor the student had encountered in his time at Georgia Tech. Some transfer students reported incidents describing instructors at other institutions.

The question format using "best" and "worst" professor was designed to elicit incidents describing a wide range of performances, not only those directly concerned with classroom work. The purpose was to delineate, as broadly as possible, the general perceptions of students as they see faculty members. Undoubtedly many such behaviors are not directly related to student attitudes or learning, or in some cases the students learn in spite of certain behaviors. However, some of the reported behaviors probably are basic to student interest in the course, attitudes toward faculty, and possibly the particular school. The best-worst polarization would seem to ensure that the student was concerned and was reporting something of importance to him. In addition, past studies have generally been limited in their coverage and no general taxonomy of faculty behaviors, as students see them, has been presented. It was thought the data collected would be a first step toward constructing such a taxonomy.

The CIT was thoroughly explained to the students early in the quarter; it is in fact a part of the required learning in the course.

The students then collected six incidents, brought them to class where they were reviewed and any deficiencies pointed out and discussed. In particular, it was necessary to discuss descriptions such as "interesting." Terms such as this were encountered quite often and present an obvious need for elucidation as to just what a given professor does that makes a course interesting. Aside from this there seemed to be no problems in collecting incidents such as availability and cooperation of interviewees or adequate time to collect incidents. In general, it appeared that the behavioral incidents obtained were accurate descriptions of faculty behaviors both the effective and the ineffective.

Once critical incidents have been obtained for an activity, the next step is to "categorize" them. Generally, this involves reading through each incident and sorting them into groups of similar incidents. Then there is a re-reading of the groups to determine whether or not the incidents have been placed in the appropriate group or category and, if needed, a re-evaluation of particular incidents. Once completed, the incidents are then placed into sub-categories under the broader formulations.

As a check on the above procedure, a random sample of 10% of the incidents is extracted from the total group of incidents before the categorization process is begun and, following the completion of categorization, these incidents are read to establish the category in which they should be placed. In effect, this is to determine the "validity" of the categories previously established.

Following the above procedure, another person reads the incidents and determines whether or not the incidents do indeed belong in the categories established. This reading is "blind" in that the incidents are numbered but the second reader does not know the category of original placement. Agreement of the two persons is calculated as a percentage of incidents placed in the same category by the two persons who have completed the independent categorizations. The procedure is a check on the "reliability" of the categorization process.

For this research, the above procedure was elaborated to insure both adequate validity and reliability of the categories established from the collected incidents.

A first categorization of the incidents was completed by a person familiar with the process, in fact, had completed a Masters thesis based upon the CIT. The classification process was found to offer somewhat unique difficulties. The first of these was that some of the incidents were actually two incidents, an eventuality not usually encountered in such work. An example of this finding:

> "the professor said he would give grades on the basis of quizzes, a term paper, and the final exam. He actually gave grades based only on the final. He gave me a lower grade than I thought I deserved but refused to talk about it with me."

Here are actually two behaviors. The first is giving students erroneous information and the second, refusing to discuss grades with students.

A second area of difficulty was the decision as to exactly which category was appropriate for a given incident. This sort of incident was where a professor discussed, or failed to discuss, how grades were to be given in the class. The question is whether the incident belongs under "Classroom Administration" or "Student Evaluation." The decision had to be made usually on the basis of the presumed intent of the student report. To this extent the categorization was somewhat subjective.

Because some of the incidents were ambiguous, the author read through the incidents, re-categorized where it appeared necessary and renamed or changed some of the categories and sub-categories.

The resulting categorization system was then given to a Georgia Tech senior student who then categorized the incidents. The two categorizations agreed on all except 39 incidents in the "effective" group, 97% agreement, and 52 in the "ineffective" group, a 96.2% agreement. The main source of difficulty in the former was with Category V, "Organization and Presentation of Material." For this Category 25 incidents were placed in other categories. With the ineffective incidents, Category VI, "Evaluation of Student Performance" accounted for about one-half of the misclassifications, most of which were changed to Category II, "Class Administration."

The incident cards were then marked and the Project Research Assistant went through the cards to check the classification system and she agreed to the extent with the devised system so that only 23 incidents changed Categories.

In general, the categorization system finally derived appears to be as satisfactory as can be constructed from the incidents at hand. The system contains six categories of effective behaviors and seven ineffective with the appropriate sub-categories under each. They are shown as Appendix 1 with the number of incidents for each.

RESEARCH RESULTS

Introduction

It was mentioned earlier that the students seemed to be reporting on faculty behaviors that were found to be quite important. To illustrate, two reports that are not incidents are presented. They illustrate the impact faculty members can have on students:

> "--shows concern for students and ability to teach his material. Has the type of personality that students can identify with personally. Students respect him as a man and as a teacher and find him affable outside of class and in class."
>
> "--relies on his national reputation as a lecturer and consultant to 'play God' in class. Egotistical, egocentric, narrow-minded, dogmatic, short-tempered, all-imposing bigot."

These subjective evaluations of professors indicate that students do react to professors in both negative and positive manners. Further, the incidents reported below show this same effect in that many incidents were reported with emotional toning to show that teachers have a quite important effect on students both in terms of student attitudes and perceived achievement.

The categories and sub-categories developed from the reported incidents are described individually below, along with illustrative incidents and some hypothesized effects on student achievement and attitudes.

Effective Behaviors

The first effective behavior category was, "Personal Relationships With Students." This general area has been mentioned fairly frequently in past studies, in particular, has emerged as a common factor in the various factor analytic studies described. However, from the data here, it appears to be much broader than previously formulated, covering behavior both inside and outside the classroom. The sub-categories are presented by number in parentheses ( ) and discussed:

(1) Knows each student by name - this behavior refers to behavior both in class (calling on students) and out of class (greeting students). To some degree it is a reflection of the machine-like nature of higher education with large classes and impersonal lecturing. testing, and grading. In such a situation, the professor cannot learn student names unless he makes a deliberate attempt to do so. As can be seen by the relatively large number of reported incidents, (12.7%), students do consider this behavior of importance; it appears to be a somewhat pathetic appeal for personal recognition. Possible effects on student achievement are questionable but student attitudes are patently affected.

(2) Mingles with students before and after class - this behavior refers to a "socializing" sort of behavior. The incidents refer to "small-talk" discussion of some news event and similar behaviors. An incident was:

> "usually comes well before the hour and sticks
> around after class to allow individual students
> to talk to him or have a small group discussion."

Generally, the category might be regarded as "fraternizing" with students and, while not related to student learning, probably has effects with regard to student attitudes toward faculty. The behaviors reported here accounted for 2.1% of the total Category.

(3) Holds social events for his students - this is probably a rather controversial behavior with regard to student-faculty relations and is rather uncommon, only 7 incidents described. The most dramatic was:

> "After the final he had a beer party that
> lasted almost all night. I found I learned
> and remembered a lot of the material much
> longer than I did in most of my other classes."

As noted in sub-category (1) above, at least some students regard closer acquaintance with faculty as important. It is possible that only the more dependent students consider this important but, at present, one can only speculate as to the personal correlates and importance of such behaviors.

(4) Gives (encourages students to ask for) advice and

assistance regarding personal problems and goals - this subcategory covers two behavioral aspects. The first is sheer availability in order to talk with students and the second, advice or help in solving personal problems. 10.8% of the Category incidents were of this type:

> "sets up appointments for each individual student to come by his office and get help on class problems and any other help that might be necessary and where he could give help."
> "I was a Junior and still undecided as to my career. The Professor who was teaching a course, not required by my major, spent a great deal of time trying to help me decide what to do. By the end of the quarter, he convinced me I should change majors and I have been happy ever since."

This behavior is exhibited not only by student request but some professors actively encourage students to come to them for help. As can be seen by the relatively small number of incidents, this is not a common faculty behavior, but is likely a quite important one with regard to both favorable student perceptions of faculty members (and learning).

(5) Discusses (answers questions about) extra-class issues with students - this is a rather rare behavior (6 incidents) as shown by faculty members and its potential for student attitudes or learning is unknown. An incident was:

> "The class could discuss current events with the Prof. This would create class interest and during the lecture the class was more interested."

Very likely the effectiveness of such behavior is related to both what is discussed and how it is discussed. As will be seen in the ineffective behaviors, students do not like rambling talks. They see them as a waste of valuable class time, even to the extent that the "professor is making a fool of himself." It appears the behavior can be effective but only when properly done.

(6) Compliments a student on a good response - this is a very rare behavior (3 incidents). All were concerned with the

professor complimenting a student for making the highest grade on a test. As will be seen, a fairly common faculty practice is to exempt high scoring students from the final examination but outspoken complimenting is almost non-existent. The actual effectiveness of compliments with regard to student behavior probably would not be of major importance but could hardly have any but positive results. Certainly, students must answer questions in a way deserving of praise, at least sometimes, but there is no indication in these data that such answers are ever evaluated. It would seem that a small effort here by faculty members could result in a more favorable learning situation, if nothing else.

(7) Explains answers in detail to all (encourages) questions asked by students - the emphasis here on "all" should be noted. This is an important behavior to students (13.8% of the incidents) and they react quite positively. Some typical incidents were:

> "Answered questions that were not really related to the course.."
> "Stopped lecture to answer my questions even though they were kind of stupid."
> "Teacher made sure student understood answer to every question, even to the point of checking with the student after class to make sure it was clear."

The general behavior is that an effort is made to answer any question that might be asked even if they are "dumb" or "stupid". Related is an active interest in making certain that the question has been adequately answered. Under the absence of this behavior, not answering questions, some of the more emotional incidents were reported. It could well be hypothesized that this behavior is the key behavior in the student-faculty relationship. This seems like an obvious behavior for teachers but some abrogate the responsibility.

(8) Treats all students fairly regardless of sex, race, etc. - this behavior in its positive aspects was not reported often (four incidents) one was:

> "Answered my questions without making some remark about dumb coeds."

Obviously, this incident was reported by a woman student and undoubtedly reflects some negative experience. The presence of this behavior is not often noticed but the obverse unfairness, is. Very likely most faculty members would regard impartial treatment of students as a given condition but partiality is probably more common than is usually assumed. The negative incidents concerning this behavior are reported under the "ineffective behaviors."

(9) Holds special problem sessions or allots class time for questions - this behavior is related to questions concerning only the subject matter of the course being taught, in contrast to category (7) above, and may be sessions other than class or part of a class hour being set aside entirely for student questions. The range of the behavior is shown by the incidents below:

> "It seemed the whole class was having trouble with the material being presented, so the Professor set up two special afternoon sessions where he would come and work problems for anyone who needed further explanation."
> "After class he would hold a private class in his office for anyone who cared to come."
> "Class before a quiz was a question and answer period. Sometimes we stayed over, once for 4 1/2 hours, until everyone understood the material."

Basically, this is the recognition by teachers that some students do have trouble learning the subject matter and need special help. The help is offered on the basis of both class and personal time, often at considerable sacrifice, to bring student learning to the highest possible level. This particular behavior has been noted or reported only in critical incidents, Douglas (1968), but the hypothesis would be of a considerable impact on student learning. Certainly the effect on student attitudes must be considerable, and it would seem a behavior to be encouraged on the part of the faculty member.

(10) Gives individual help, in class or office, without hesitation (encourages students to ask for) - this is by far the largest sub-category, 41.4% of the incidents. In contrast to the category above, it is concerned with encouraging students to ask for or offering individual help. Some reported incidents were:

> "The instructor helped me on a computer program after class. He seemed to enjoy helping me as long as he knew I was trying."
> "Would come to the campus at night or on weekends to help, whenever it was convenient for the student."

> "He made sure your questions were answered when you went in for an office visit. He not only worked the problem I was having trouble with but he had me work two or three similar problems to make sure I understood. He also worked these problems in an orderly, clearly written manner on a piece of paper which I could carry with me."

This is a rather complex behavior which seems to involve several facets. One is *willingness* to help. Students apparently sometimes receive help but it is given grudgingly in an ungracious manner that they feel is demeaning. Availability for conference is another facet, that is, the professor is "always in his office" or makes and keeps individual appointments. There is also the facet of determining that students do, in fact, understand the material and, finally, a willingness to meet at odd or inconvenient times to give individual help. The entire complex of behavior indicates a sincere interest in student learning and a willingness to make every effort to ensure that they do learn. Again, it could be hypothesized that this behavior, willing and individual help, would have a sizeable effect on student learning. In fact, this may be a key behavior in differentiating between the effective and ineffective teacher. The same effect would probably be found in terms of student interest in and attitudes toward the course.

(11) Miscellaneous - a small number of incidents (15) were reported that seemed to impress students favorably but do not fit well into any of the ten sub-categories reported above. Some of these were:

> "Prof's criticism of students work led 3 out of 5 student contestants to win or place in a national award contest."
> "Held class at his house in a casual atmosphere."
> "Took an interest in student. Stopped me in the hall and asked how I was doing on a project."
> "Never told a co-ed joke" (obviously reported by a woman).

The importance of behaviors such as the above is difficult to estimate. They tend to indicate a real interest in students and a rather unconventional approach to teaching; often these are related only to an individual relationship Probably they have an impact on individual students but their effect on either individual or group behaviors is a moot question.

The entire category indicates that student-professor person relationship can vary over a tremendous range of behaviors. The underlying determinant would appear to be a sincere interest in student learning and welfare on the part of the individual professor. This is shown by active encouragement on the part of individual professors that induces questions or appeals for help. This not only is related to actual class work but extends to personal problems ranging from relatively trivial things such as loaning students small sums of money to important effects on course understanding and learning.

This general area of behavior was noted in some of the factor analytic studies under some rubric such as "empathy". The behavioral area, from the results here, appears to be real enough in terms of total incidents reported, 29.8%, and is remarked by students. "Socializing" with students, treating student questions with respect, involvement in both personal and learning problems seem to be of real importance from a common-sense point of view. In particular, special problem sessions with individuals or an entire class would indicate that such students learn more and have more interest in the subject matter and positive attitudes toward the course. Or, generally, it would appear that professors showing the behaviors comprising this first category would interest and motivate students so that only positive results could be expected, especially, when the negative behaviors, to be described later, are encountered as a contrast to the positive.

The second effective behavior Category is entitled "Classroom Administration." The sub-categories are described below:

(1) Extends time limit (change dates) on assignments and quizzes - this behavior refers to extension of time or date changes largely because of some contingency condition as:

> "He knew you had other courses and gave you plenty of time to do projects - he once gave the entire class an extension on a project because everyone had a math quiz."

This behavior was infrequently reported (6 incidents) possibly because it is infrequent. The indication is both of an appreciation of student work loads and a flexibility with regard to his own course requirements. In effect, there is a recognition of priorities with regard to other course requirements, student stress, and personal course requirements. The more flexible person is willing to change plans and requirements to adjust to unforeseen contingency factors. Very likely this has positive effects on both achievement and attitudes.

(2) Informs class of days he will be absent or changes in plans - again a very small category (4 incidents) possibly because few faculty members exhibit such behaviors. One incident was:

> "Scheduled a test for one day and later changed his mind. He then sent a letter to each student and told them the test was put off another period."

Possibly one reason why this behavior was so infrequently reported is shown by the incident above; very few people are likely to make such an effort. Obviously, there would be occasions when it would be impossible to let students know of changes in advance but there seems to be little effort to do so at any time. With such infrequent occurences, it would be difficult to determine the effects of the behaviors in this category on students. It would be hypothesized that they would be minor.

(3) Lecture begins and ends on time - again a quite infrequent incident (4) possibly because it is a rare professor who is described in this way:

> "The Prof was very punctual. He was always right on time to class and would dismiss class on the whistle (a Georgia Tech institution-author) regardless of what he was doing at the time."

(4) Distributes or details a study plan the first week of class outlining the course requirements - many incidents here are dual covering this and the next category. A single incident:

> "Prof gave out all of the assignments on the first day of class so that we knew exactly what we had to do for the rest of the quarter."

Only 11 incidents (9%) were reported in this category. The behavior seems to be more uncommon than one might suppose or students regard it as routine and do not report it, probably the former. No doubt this assists student achievement.

(5) Follows course syllabus or lecture outline as scheduled - here ten incidents were reported, most of them of the dual nature:

> "Work was covered exactly on schedule as it was outlined in the course outline handed out at the beginning of the quarter."

The two categories (4 and 5) together constitute a sizeable percentage (18%) of the incidents in the Category. Strangely, lack of this behavior receives scant attention under the reported ineffective behaviors. Some students apparently regard a detailed course description and adherence to its schedule as an effective behavior but absence of these behaviors is not remarked. It would be hypothesized that this behavior on the part of faculty would have relatively important effects, particularly, on student achievement.

(6) Gives examples of quiz items or what to expect on quiz in class - again a small category (6 incidents), however, students do seem to appreciate the effort made. Again the small number of incidents is probably due to the rarity of this behavior. No incidents are shown because they contain the obvious.

It is interesting to speculate what effect this behavior might have on students. The hypothesis would be that better grades on quizzes could be expected, in particular, where sample questions have been solved in class.

(7) Keeps old quiz questions on file for student inspection - again a small category (10 incidents). The comments for the above sub-category apply here.

(8) Requires and grades homework - one of the larger subcategories here (26.2%). This behavior of requiring homework is seen as both effective and ineffective, however, graded homework is generally seen as important in helping student learning. Two such incidents were:

> "Homework was assigned and always taken up and graded. The Prof always went over the problems assigned to make sure the students understood what they were doing."
> "He required homework to be turned in, thus, forcing the students to learn the material."

Potentially, if homework is assigned, graded, and discussed, it would indicate student learning and general class progress. From the number of incidents reported this is a fairly common teaching device but, as indicated, receives a mixed reception from students. An interesting study might be to compare the groups of students who have positive and negative reactions to such assignments.

(9) Grades papers or quizzes promptly - again there are relatively few incidents (7) and all deal with returning work no more than two class periods after it has been collected. The few incidents are due to the rarity of this behavior, as shown by the same sub-category under the ineffective behaviors.

It is questionable that this behavior would have a real effect on student learning but probably would be found to have a rather marked effect on attitudes toward the teacher.

This general area of teacher behavior has been remarked in the research previously described. Various descriptions such as "efficient", "business-like", and so on have been the terms used. The behaviors seem to indicate an organized and planned approach to the course but, at the same time, flexible with regard to student needs. That is, the organized and planned approach is the more desirable but is not a major point; the student is. Students seem to notice the attempt at organization and appreciate it as shown by the reported incidents. It can probably be safely supposed that, in relation to the behaviors shown here, student reactions would be positive in contrast to those exhibiting a characteristically slip-shod approach.

The next category, "Student Participation" has received much lip-service but is generally so ill-defined as to be almost meaningless. The behaviors described below seem to give some semblance of an operational definition.

(1) Permits students to determine part or all of course content, class policy - the behavior here is to ask students what they want to study or discuss and includes votes on class policies. For example:

> "The Professor ran his class democratically. When a question arose as to a day for the mid-term, he gave us four days to choose and we voted."
> "Realizing his course was the last in a series, the Prof allowed us to choose a topic to discuss for the quarter which would require application of what we had learned in basic courses. It was my most interesting course."

The use of the participative techniques implied above is limited to some degree by the course, but the use of democratic procedures is not. A fairly substantial percentage of incidents (16.1%) indicates that students do appreciate the participation. Whether or not it makes learning more effective is undetermined, but students did report heightened levels of interest under such participative procedure.

(2) Improves his course by making changes based on criticism requested from students - this is relatively small category (9 incidents) probably because it is so rarely done as:

> "--asked the class after three weeks what we liked or disliked about the class, the material, and himself. After discussing this for two class periods the Prof used our ideas and actually changed his presentation of the material, chose material to be discussed that was the most interesting, and made the class one of the most interesting I have ever had."

It is unlikely that many professors would go to the lengths illustrated by the incident above; in fact, it seems to be a rare occurence that students are ever asked to directly evaluate a given teacher's performance. In the case above it seemed to be quite effective, at least in terms of interest, if not in terms of actual learning. This critical function for self-improvement is likely to be the major utility of any teacher evaluation device. Probably any teacher could benefit to at least some degree through opinion and criticisms. As has been shown, both teachers and students tend to misperceive the other and this is a possible tool of correcting such misperceptions.

(3) Schedules quiz at convenience of class majority - this is a relatively small category (11 incidents) probably because of the rarity of this behavior on the part of faculty. An illustrative incident:

> "Realized students were taking more than one subject. Always asked class when they wanted quiz."

The general point of these incidents is that the nature of academic scheduling tends to make examinations, term papers, etc. fall on or near the same date. It is possible for students to have as many as four examinations scheduled for the same day.

At the same time, these are in no sense absolute requirements and some flexibility by faculty seems indicated. However, judged by the number of incidents, this seemingly simple accomodation is uncommon, although it is appreciated by students and probably affects performance (study time).

(4) Encourages group discussion, questions, and differences of opinion - this is by far the largest sub-category here, (52.4%) of the incidents. Apparently, this is a characteristic manner of conducting class on the part of certain faculty members and not others. Some typical incidents:

> "Never lectured; carried on Socratic method of dialogue."
> "Class was small group discussions. The teacher was more-or-less the chairman. We learned more because of this."
> "He held class discussion every day and got everyone to participate by asking individuals questions."
> "He would assume unlikely reasons for events instead of the likely ones. Then asked students to justify such reasons. We had to think."
> "Prof invited students to question his ideas and theories. Did not hold it against them if they had opposite ideas."

As can be seen, there are a variety of behaviors that can be used to induce student class participation and, from the number of incidents obtained, which seem to be seen as effective by students. In particular, not allowing disagreement to affect teacher opinions of students seems to be a key factor. As will be seen, under ineffective behaviors, this is not true of all faculty members. Participation seems to raise student interest in the course and, at least some believe they learn more. However, the technique would seem to be limited by the subject matter under consideration. Inviting questions and personal experience might be appropriate in some cases but, in the case of highly technical subject matter, only invited questions are likely to be relevant. Another point is that some persons are likely to be more effective using participation than others. In general, while students seem to appreciate its usage, its effect on student achievement needs to be assessed.

(5) Seeks feedback from students, in particular, understanding of material - the distinction of this sub-category from (4) above is the concentration on course material. Specific questions are asked either on the material *per se* or generally as to understanding of the course. The incidents are of an obvious nature and constitute 15.3% of the Category. The relationship of this teacher behavior to student learning seems apparent and very likely has positive effects on student motivation. The behavior seems to be one that could be adopted for any class and for any subject matter.

The fourth behavior Category constructed was "Classroom Presence." It covers a somewhat heterogeneous set of behaviors and seems largely to involve both a knowledge of subject matter and a concern with personal characteristics in its presentation. Possibly is related to the "annoying habits" found in the study by Moore (1937) *op. cit.*, as described earlier. This sub-category is a rather small one, 5.1% of the incidents, probably because the positive behaviors shown are not as remarkable as the negative, since the positive ones are largely taken for granted by students.

(1) Makes dramatic gestures (comments) to emphasize important points - this was a very small sub-category (3) incidents and apparently is not characteristic of faculty behaviors or, possibly, is just not remarked by students unless the behavior is extreme as:

> "Enlivened the class by dancing around to
> demonstrate how molecules bounce."

Even though such behavior seems rather rare, it would make an interesting study as to its effects as a mnemonic device, and probably would result in longer retention of at least the one point.

(2) Speaks in a clear, distinct manner; correct diction or grammar or both - again a small category (3 incidents) probably because the positive behavior is expected whereas the negative is of importance and more often remarked.

(3) Uses humor that stimulated class interest and attendance - the largest sub-category (58.4%) of these incidents. Some incidents:

> "Always had a joke or off-beat comment to
> begin class. Put class at ease."
> "Interspersed his lecture with humorous
> stories from his own life that illustrated
> the point he was trying to make."

"Got and held class attention by jokes
about the subject."
"Once when writing on the board, turned
to a visiting Prof and said, Hell, I'm not
really sure about this but I think it's right."

The number of reported incidents indicate that students appreciate the use of humor in lectures. The humor seems to serve both to clarify and dramatize material and make the professor more "human" by inducing a relaxed and informal class atmosphere. Very likely, appropriate use of humor could be quite effective both in terms of student learning and teaching particularly if used to stress important points. Obviously, it would be a difficult variable to assess in terms of prediction of student achievement; however, the large number of positive incidents reported seem sufficient evidence of its importance to student learning and interest.

(4) Lectures without relying solely on (reading) notes or text - only two incidents were reported here and are included to show the positive side of what is a large ineffective sub-category. Again it would seem the positive behavior is expected and not remarkable by students, whereas the negative is viewed as important.

(5) Uses language that students can understand (not talk-down to) - the incidents reported here are not very well defined. For the most part, they say that the material is "translated" into "terms students can understand." This seems to be an important feature of lecturing, (21.5%) of the incidents but is quite difficult to illustrate with appropriate behaviors. This is an effective behavior that likely would be recognized by students when it occured but is difficult to specify, because it tends to be material-oriented, that is, the specific behavior would differ by subject matter. In judging, it is also probable that there would be some variability among students with regard to those who could understand and those who could not. It would appear to be an item worth including in any teacher evaluation device but needs further elucidation to clarify and objectify it.

(6) Personal appearance - again a small category probably because the positive behavior is the norm, that is, most professors make a presentable appearance. In terms of student learning and interest is probably a rather trivial consideration but could have some effect on student attitudes toward a particular faculty member.

The fifth Category, "Organization and Presentation of Material" is by far the most commonly reported of the effective incidents, 32.0% of the total. Again, this area of behavior has been shown to be of key importance in both the rating and factors analytic studies previously reported, in fact, in one case, Solomon (1966), *op. cit.*, it was shown to be the *only* behavior related to student achievement. It seems well within the realm of possibility that studies of effective teaching ultimately could be concentrated in this area along with other behaviors desirable but not critical with regard to student learning. This Category contained a wide variety of behaviors, 19 sub-categories. Proper evaluation of the reported behaviors will require a somewhat lengthy form until experimental evaluation can determine which of the behaviors are the more salient.

(1) Begins each class with a review of previous work as:

> "The Prof gave a 5 minute summary at the beginning of each class meeting of what was covered in the last period."

Only six incidents were reported here probably because this is a relatively uncommon faculty behavior, the assumption being that students remember the last class topic of discussion. Whether or not such an assumption is warranted is questionable and the practice would seem to be conductive to student learning and understanding. Such an hypothesis is susceptible to experimental test and the actual value of reviewing will not be known until such tests have been accomplished.

(2) Stresses important points, general concepts in teaching - here the reported behaviors named are to some degree specific to the particular subject matter; however, the general behavior was to intentionally draw student attention to the basic or difficult ideas and concepts (often the connotation was the material would be included in examinations). Some incidents were:

> "The best Prof this student knew presented lecture material in outline form on the board with major topic headings to aid note-taking."
> "Professor states what is important (exact subjects) before beginning lecture."
> "Explained important parts of the text and told us what mistakes we were likely to make *before* we made them."
> "Prof told students exactly which equations in the

text were necessary to know and which ones unnecessary."

This category is relatively small (5.0% of incidents) but the behavior could be hypothesized as quite important to student learning. Certainly from one's own experience, it is evident that knowledge of what to learn would enable concentration on these aspects and thereby save time and effort. Also it would by hypothesized that such teacher behavior, consistently exhibited, would result in superior student achievement.

(3) Puts important information on the board in a clear concise manner - it will be noted here that these are two behaviors. Both the information and its legibility are involved. On the ineffective side, a student complaint is that often blackboard presentations cannot be understood.

> "He wrote all definitions on the board - making difficult material easier to learn."
> "Wrote main points on the board to be sure we understood them."
> "Every note he expected us to know he put on the board."
> "His equations and formulas were written very plainly and orderly on the board. He also left them up there as long as possible."

This sub-category is related to (2) above and is a technique for stressing important points and concepts. The comments above apply to this behavior.

(4) Uses current and pertinent examples and illustrations to explain material - this appears to be a somewhat important behavior (9.7% of category incidents) with regard to students seeing a professor as effective. The specific behaviors reported are quite heterogeneous but the general theme is to aid understanding of the subject matter. Most of the incidents were non-specific in that they described habitual behavior of tying material to current events, presenting simplified examples, or showing relevance to a particular field. However, some were quite specific as:

> "For example, relating Spanish culture to present (U.S.) events. Spanish influence on present folk-rock music."

Others, less specific, were:

> "Current events introduced, especially as affecting companies in this area."
> "Although this was just a survey course for non-majors, the Prof showed how the subject was relevant in other fields. He tailored experiments to the non-majors."

This seems to be a fairly common teaching practice according to the number of incidents reported. Again, examples and incidents would need to be specific to particular subject matter but students seem to find the behavior of importance, in particular, to clarify difficult points. The probable relation to student learning seems obvious.

(5) Shows relevance of material to the "real world, "the student's major, and/or student's outside interests or future - this is by far the largest sub-category (26.0%). The particular behaviors seem to be an effort to arouse student interest by specifically pointing out the relevance of the classroom material to later life. Some incidents were:

> "He would introduce the theory behind a problem and work several examples (not one or two) of the problem and its applications."
> "He would often discard material in the text saying it was not applicable in the real world and we would never use it."
> "In a course on project management, he used actual examples of real projects in which he had worked as a consultant and showed how the methods taught in the class applied to these projects."
> "We spend 3 class meetings discussing (course) application to campus riots."

The behaviors here are varied in that a professor shows practical application of theories (as equations), verbally relating course topics to the "real" or future world or, most commonly, using his personal practical experience to show relevance of the classroom material. Comments to the sub-category above apply here.

(6) Asks questions in class; stimulates thought, interest - this ancient teaching device, one student mentioned the Socratic method, needs little comment here except that it does not seem to be common (4.8% of the category incidents.) Where it is used, students seem to find it a refreshing change ("not just talk and write on the board") and its effectiveness for student learning, interest, and attitudes needs to be evaluated.

(7) Admits answer he does not know/provides answer next lecture - only seven incidents were reported here but, on the negative side, this seems to be a quite emotional issue with students, especially, where they see professors as "making-up" answers. Some incidents were:

> "Doesn't try to make up an answer to a question if he doesn't know the answer; just says he doesn't know."
> "The Professor did not beat around the bush in answering questions. If he didn't know the answer, he would say so. He would then find it and tell us next period."

As far as this behavior is concerned, the author knows of no effort to assess its resultants in terms of student behaviors. It might by hypothesized that students "test" professors by asking questions and certainly respect for the professor as a person might hinge on the perceived results, but any behavioral correlates of this behavior remain to be determined.

(8) Lectures reinforce the textbook - here only five incidents were reported and were again rather general statements of habitual behavior. This behavior would be difficult to make objective in specific terms except as asking whether lectures closely followed the text or were presented as original material. Consequences for student achievement of this behavior are likely to be tied quite closely to the nature of the achievement measure used. For example, with tests involving largely factual material, the effects are likely to be minimal but, with achievement measures stressing conceptual and integrative learning, major differences in student achievement might be hypothesized.

(9) Supplements course (book) by using outside reference materials - the behavior here means actually bringing in and using outside material or referring students to such appropriate materials.

A fairly substantial sub-category, (5.8% of the incidents). Some of these were:

> "Suggested other books to read to help you understand the material."
> "One day he described for us several technical magazines related to the course and offered to help us obtain issues we were interested in."
> "Would bring Fortune magazine to class and talk about the articles."

The author would hypothesize that this behavior might be one of the more important in discriminating on an effective - ineffective teacher dichotomy. It seems to indicate a teacher is knowledgeable in the newer literature. The inference being he makes a conscientious and continuing effort to make his class more interesting. Effects on student behaviors would be hypothesized quite sizeable as a result of such behavior.

(10) Distributes hand-outs and/or copy of class notes to supplement course (quiz)- this seems to be a fairly common but by no means universal effective teacher behavior. Illustrative incidents were:

> "The Professor would always bring Xeroxed examples and explanations from other sources of complex problems to class."
> "He would provide hand-outs on all his lectures. It helped students pay more attention knowing he didn't have to take down everything the Prof said."
> "Prof passed out a Xeroxed copy of his notes each day."

The behaviors described here seem so directly related to student learning that it is remarkable that it is not found to be more common. The functions are both to indicate important points and to inducing increased student attention by eliminating the need for notetaking. Any empirical test of this behavior would hypothesize benign effects on all student behaviors.

(11) Supplements lectures with visual aids (black-board) - the behaviors here are contrasted with those in category 3, above, in that only material of basic importance is put on the board by way of special emphasis. Other visual aids have this same connotation,

although this may not always be the case. Some incidents were:

> "Always put important ideas on board."
> "The Prof illustrated his lectures with slides he had accumulated his trips abroad."
> "He used films and tapes to add to understanding of the material and give a wider view of the course."
> "I enjoyed this Prof's class because he brought in working models that demonstrated and clarified the theory we were studying."

This behavior, of course, is limited by the availability of appropriate aids. Studies that have been done to evaluate the effects of such behaviors have indicated heightened student interest but only minor effects on achievement. Possibly more relevant criterion measures might show such effects.

(12) Provides field trips - a very small category (4 incidents). This time - honored teaching practice needs no comment here.

(13) Invites guest lecturers who are specialists on course topic - again a small category (6 incidents). Probably most would agree that effectiveness is highly correlated with quality of the guest.

(14) Explains (works out) answers to quiz, homework, and class problems - the behavior here is to determine that students know the answers to all problems presented. As shown by the following incidents there are various techniques that can be invoked:

> "Prof assigned homework and on the day it was due he would hand out answer sheets to the homework. He also handed out answer sheets for quizzes the period following the quiz."
> "The Prof spends at least half the period working examples."
> "The Prof would explain the material once then work various types of examples with material."
> "He goes over what to do then lets you do some problems in class while he can give help and answer any questions you have."
> "Would take time to go over mistakes made on a quiz with students."

This behavior is another that seems habitual with some professors, a teaching method. The key point is that no wrong answers are left uncorrected, and from the viewpoint of learning theory, quite a significant approach. In general, time is devoted to student learning rather than teaching, assuming the distinction is clear. The presumed effects of this behavior on student achievement would be of major importance.

(15) Does not regularly follow book or notes in his lectures (well-organized and prepared) - the inference from the behavior, in parenthesis, is one students seem to draw from the behaviors reported (9 incidents). Some incidents:

> "Disregarded book in lectures."
> "Never used notes but stayed on the subject."
> "Prof would branch off but return exactly to the topic discussed."

This behavior does not seem to be the product of rote memory or sheer familiarity but indicates a thorough preparation before the lecture is started. It seems to inspire confidence in students and could possibly be related to student learning; no real empirical evidence is as yet available on the possibility.

(16) Uses department (personal) experiments, projects, or work to stimulate student interest - the behavior is a fairly limited one (4.6% of category incidents) and has to do with relating course material to current activities, usually research. The primary intent seems to be to arouse student interest and, from the incidents, is effective doing so. It was not mentioned, but this behavior might well be effective in motivating students, especially majors in the subject. It would even be possible to prepare formal descriptions of on-going work for distribution to students in order to stimulate and motivate. The hypothesis would be of enhanced interest and achievement.

(17) Has full (or personal) command of subject matter - this is a rather difficult evaluation for students to make but they do so on the basis of behaviors such as:

> "Understands material so well he doesn't even bring notes to class."
> "Knew material - could always answer any question."
> "Knew subject so well did not need to derive complex equations."

One could quarrel with such behaviors as indications of knowledge but students do make these inferences. Testing them would not be overly difficult even though their relation to student behaviors might remain nebulous. The subject matter knowledge-teaching effectiveness relationship would be a particularly interesting subject to study. Such results as exist now are both scanty and contradictory.

(18) Reviews material before a quiz or assignment (study guides, assigns similar problems) - this is a small category (11 incidents). Probably this is because the behaviors are a deliberate allotment of class time to problem review sessions and prepared materials similar to those to be encountered as:

> "--before tests he gave out some problems he said would be similar to test. If we knew them we could pass the test."
> "The best Prof this student ever had held two days of well-organized review prior to finals during which he went over all the problems in the course...thanks to the Prof's effort the grades were good."

Such behavior requires extra effort on the part of faculty members but probably results in better student achievement. What is not wanted is deliberate "coaching" for a specific test but proferred guidance certainly seems a desirable and quite probably an effective behavior from the student point-of-view. There would seem to be every reason to encourage such behavior by teachers.

(19) Pace of lecture can be followed (to take notes) - this behavior with relation to students seems quite apparent. It is a deliberate effort by a professor to keep his lecture pace at a rate which allows students both to follow the material and take appropriate notes. Only six incidents were reported; it is possible that only a small sub-group of students require the noted slower lecture pace.

The sixth and final behavioral Category, "Evaluation of Student Performance," is one which possesses the most emotional potential. Since effects of faculty behaviors here have direct and immediate effects on students, they understandably are quite concerned. The concern takes the form of knowing how evaluations are made, the methods used, and grading adequacy and fairness. In addition, there is the use of such results and professor willingness to examine his own evaluation practices. The importance of this category by students might be inferred from the fact that these incidents are 17.3% of the total effective incidents. In addition, it might be well for teachers to realize that as they object to unqualified or unfair evaluations of their performance, students have the same feelings. A point that should be made here, as is also true of category 5 above, is that many of the behaviors reported are dual incidents, that is two behaviors reported. Individual professors, if effective in the given area, seem to generalize their behavior in the area to several component behaviors.

To at least some extent whether or not this is true of areas other than evaluation of performance is not known and can only be determined through empirical use and evaluation of a developed evaluation device.

(1) Tests based upon lectures, text, and/or homework (announced and relevant) - this is by far the largest sub-category (31.5% of the incidents.) Some typical incidents.

> "Gave quizzes from notes handed out."
> "Tests were always on material assigned or covered in class."
> "Outlined what he expected us to know on a test and then tested us on this material rather than throwing in something we had never seen."
> "He was always telling us things to remember because they would be on the test. Ninety percent of the test questions were taken from this material."
> "He works the exact type of test problems that are on his tests in class."
> "Gave specific material on which tests would be given. Was a small enough range of material that one could get a good grasp of it."
> "When tested on outside material, Prof told us specifically what material would be covered."

If performance evaluation is to take place, it seems an obvious requirement that a student should be tested only on material he could be expected to know. However, as seen here in the number of reported incidents (and the later ineffective reports), the behavior brings out student comment indicating effective teacher behavior. The basic requirement of any performance measurement is relevance and here students are by implication saying that this is so rare as to bring out special comment when it occurs. As any faculty member knows, student complaints about tests are quite common, and usually shrugged off. The data presented here indicate that the student complaints may have some real basis in fact and are not just defense mechanisms. A consideration here might be some faculty training in psychometric principles or, possibly, professionally constructed tests. There is little doubt that this area is of serious importance to students and warrants careful thought.

(2) Tests require knowledge of principles rather than memory alone - only six incidents reported here possibly because the behavior is so rare, that is, construction of an appropriate test as:

> "This Prof gave a quiz that made you apply principles learned in his class; not just regurgitate his lectures."
> "Quiz questions never came directly from the book - they required the student to use reasoning rather than memorization."

Construction of tests as indicated above is difficult, particularly in the more abstruse subjects, but the point is use of material learned. This, in fact, might serve as a definition of learning but apparently is not always tested. Previous comments about test construction apply.

(3) Gives take-home final and/or open book quiz (use of class notes) - the effectiveness of this type of evaluation is likely to be a point of sharp controversy. It is an open invitation to cheat but is also an opportunity for students to show their mastery of the material. The incidents offered were obvious but the question as to the desirability of the technique is not. Its effectiveness as a teaching and evaluation device seems fraught with all sorts of difficulties and any definite answer appears unlikely.

(4) Schedules quizzes at regular intervals - only eight incidents were reported here and were seen as effective in terms of both "keeping up" and "knowing where you stand." Mostly the quizzes were given weekly and again this technique could be controversial. On the one hand they do allow self-evaluation but on the other are a constant threat. The hypothesis would be here that certain student characteristics are the determinant of preferring or not preferring this practice.

(5) Writes comments (reviews) on returned papers and quizzes - again a small sub-category (7 incidents) with obvious behaviors. The comments for the sub-category above apply here.

(6) Students with high average are excused from the final exam - again the incidents presented here are obvious and a small category (11 incidents) probably because this is a somewhat unusual behavior. In terms of the incentive theory of learning this is a quite appropriate behavior, in particular, if students attaining above a certain announced level of attainment would be affected rather than one or a few students. Experimental tests of this behavior could be rather easily arranged.

(7) Students with low average are permitted to do extra work (test) - again a small sub-category (13 incidents) with obvious behaviors reported. This practice is appreciated but is not defensible in terms of student evaluation in that certain members of the class are evaluated on a different basis than others. If all are given the opportunity to improve their grades, the situation is different but there is no indication of this in the incidents. An equitable evaluation system requires the same standards applied to all participants.

(8) Disregards the lowest test score of each student (optional tests) - a small category (11 incidents) but with varied behaviors as:

> "Went into final failing, made an A on the final and Prof dropped the quizzes and gave me an A."
> "Prof gave us a choice on whether to take the mid-term; if we took it, he would only count it if we made a C or better."
> "Prof gave an optional 3rd test which would be averaged into your final grade if it would help you and thrown away if it would hurt you."
> "Always dropped lowest quiz grade."

The effects of this optional grading system on student behavior have never been assessed to the author's knowledge. Any hypothesized findings appear to be the sheerest conjecture.

(9) Permits make-up tests at individual convenience - a fairly uncommon report (13 incidents), and again with quite obvious incidents. The behaviors are first, allowing make-ups, and

accepting student excuses (believing him), or just accepting a reason for missing tests that is not officially sanctioned. The students did not report their reactions to this sort of behavior on the part of professors; it would be hypothesized that the effects would be minor.

(10) Takes into account class participation, application, and/or effort in assigning final grades - a wide variety of faculty behaviors are shown in this sub-category as:

> "If you do badly on a quiz he will ignore it
> if you have shown interest and participated
> in class discussions."
> "Prof counted outside work enough so that
> if you did it you would pass the course."
> "Based grades solely on class participation.
> Everyone who contributed in a meaningful
> way got an A."
> "Told me I got a D instead of an F because
> I showed an interest in class and tried hard."

Here if some measure of "interest", "participation", or "effort", is available, such practices can be defended. However, there is opportunity for subjective, even selective, grading even though students see this as effective behavior. The hypothesis would be that certain student characteristics would determine whether or not such behavior was seen as effective.

(11) Curves grades on the basis of class distribution - this is a fairly large sub-category (9.4%) and the incidents given are obvious and related to relative rather than absolute evaluation of test scores. Although this is a recommended psychometric practice it is uncommon, as will be shown in the ineffective incidents. Actually, there is no justification for not having grades determined by a scaling practice of some sort. Where a large proportion of the class is receiving low or failing grades as a result of some absolute standards, the difficulty probably is not attributable to the students.

(12) Does not penalize for class absence or tardiness (accepts excuses) - this behavior by professors, not requiring attendance, is seen as effective by students mainly with relation to accepting excuses that are not officially sanctioned. Mentioned here is "trusting students." The incidents are obvious again but there seems to underlie a resentment of required class attendance. There are only indirect allusions as "never took roll" or "didn't

count cuts", but this seems to be why students see the behaviors as effective. It would be hypothesized that student attitudes toward such courses would be favorable but achievement poorer, in particular, if the achievement criterion was a test of a conceptual or integrated nature.

(13) Reviews test scores and changes grade if warranted - here only six incidents were mentioned. It is difficult to believe such faculty behavior is so rare. Possibly what is rare is students asking for a grade review as:

> "I went to a Prof about a grade change. He did not have my mid-term grade of 100 recorded. When I told him the grade, he took my word for it and changed my grade from C to B."
> "Prof was always willing to discuss grades."
> "Volunteered to review tests in relation to grade in course."

It would appear that willingness to a least review a student grade is a fundamental faculty responsibility. Grade change, of course, would depend on the results of the review. The impact of this faculty performance does not seem apparent as related to student behavior.

(14) Grades papers himself rather than employing a student grader - only four incidents, of an obvious kind, were reported. Probably because most students do not know how their papers are graded. Again, effects on student behavior are not readily apparent.

(15) Adequate time to complete tests - a very small subcategory (1 incident) but important as will be seen from the ineffective behaviors. The deleterious effects on student performance of failure to allow adequate time to complete tests can be imagined; however, those in a positive direction are somewhat questionable. It would be hypothesized that both achievement and attitude toward the course would be better where adequate time is allowed for test completion.

(16) Miscellaneous - only five incidents were reported as:

> "Graded on the basis of student improvement."
> "Pop quizzes forced me to study."
> "Gave me a higher grade than I deserved."

One can only speculate on how "student improvement" was assessed.

The effective faculty behaviors presented above have covered a wide range of heterogeneous behaviors. Some experimental hypotheses, with possible results, have been suggested. Full integration of these results will be attempted following discussion of the ineffective behaviors in the next section.

Ineffective Behaviors

As will be seen, many of the ineffective behaviors are the obverse of the effective behaviors discussed above. However, there are exceptions in both categories in that only effective or ineffective behaviors were reported. In the evaluative questions shown later, only one item is required to evaluate those behaviors which have an obverse whereas some specific behaviors require their own descriptive item.

As with the effective behaviors, the first behavioral Category is, "Personal Relationships With Students", and again the sub-categories are presented following their number designation in parentheses.

The first area of complaint is one that probably most people have entered with some teacher. The incidents reported represent some 5% of the Category total. Possibly the behavior is not common but it does infuriate students as shown by the language of the reported incidents (edited out for this report).

(1) Shows favoritism toward some students (athletes, "apple-polishers", reciters, etc.) Some incidents were:

> "Picked leaders for class projects and told the class these were the smartest students."
> "Gave very simple quizzes that everyone did well on. Gave all athletes A's when they did not deserve it."
> "--required class answers and statements throughout the quarter.---student studied and tried to be called on but the Prof always overlooked him. Finally, he started raising his hand on the hard, easy, and all questions. He was never called on. He went to see the

Prof and he told the student he would remember to call on him more but it hardly improved at all. At the end of the quarter the grades for most students were above his. Talked to the Prof about it and was told had not marked his answers to the list of questions."

The effects of this sort of behavior on student achievement have not been tested but it would be hypothesized that motivation would be seriously effected and, along with it, achievement.

From the incidents it is apparent that this sort of behavior is quite easily observed by students, probably more than most teachers would believe or accept. Aside from achievement, it would be assumed that students would lose all respect for such a professor including those who benefit by the blatant favoritism.

(2) Singles out some students as inferior (discriminates) - the behavior described here is where the students are told they are "inferior" or in some way such indication is made for an apparently illegitimate reason. The behavior is not commonly reported, about 5% of the Category incidents, as:

> "Made fun of anyone who made a low grade on a quiz."
> "Kept asking this one student questions he didn't think he could answer."
> "Told me I would fail no matter what I do in the course."
> "--told me he couldn't give me an assistantship because I was a woman and men were more deserving. I had to quit school and work long enough to come back and pay for it myself."

The results of this behavior are probably quite similar for those described in (1) above where the behavior is apparent to the entire class. In the case of single individuals it is likely to be quite demoralizing and generally result in adverse attitudes on both an individual and group basis. An hypothesis would be some personality defect in faculty members who exhibit such behavior.

(3) Ridicules or embarasses students (question or performance) - this behavior is somewhat similar to those in the sub-categories above but refers for the most part to single incidents and

characteristic behavior not directed at any particular individual or group of people. It appears to be a fairly common behavioral event, 22.4% of the Category incidents. Some incidents were:

> "Anytime someone asked a question he said, if you had read the book you wouldn't need to ask that."
> "Ridiculed a girl in class until she cried."
> "I asked what I thought was a legitimate question and the Prof said he couldn't stop to answer it."
> "--made fun of students if they gave unsatisfactory answers to questions."
> "Refused to answer any questions by saying that only stupid people asked them."
> "Criticized my paper in front of the entire class."
> "If you ask a question the Prof says, do you really need that answered, how did you get this far at Tech."

Granted that students do ask "dumb" or "stupid" questions, it is necessary to recognize that they do not often do so deliberately. It is also necessary to recognize that some course material may be difficult for an individual student even though most students understand it. Since the presumed intent of teaching is to have students learn there seems to be no adequate defense for the behavior described above. In terms of student achievement, such behavior can hardly have any but adverse effects and, from the tone of the reported incidents, results in contemptuous attitudes toward the teacher involved. It would be hypothesized that any experimental test of such behavior would show rather sizeable effects on both student attitudes and achievement.

(4) Loses control of emotions in dealing with students (shouts, curses, etc.) - the behaviors reported here refer only to those exhibited in the classroom. Presumably this occurs in other contexts but it was not so reported in the incidents collected. This is a fairly large sub-category, some 8% of the incidents collected, as:

> "Often cussed out the class-soundly-for reasons no one could see."
> "One student, in a humorous way, made light of the fact that the Professor made

> an error on the board. The instructor became angry and dismissed the entire class."
> "--would make us sit in assigned seats and became very upset if people didn't sit in them. Got so emotional once, walked out of class."

Again the language of the incidents has been edited somewhat because of the emotional nature of students' responses to the behavior above. The general tone is probably best described as one of disgust at such behavior. It seemingly violates the "role expectancies" by students and is quite important to them. Probably would have relatively little effect on student achievement but would be a major determinant of student attitudes.

(5) Harasses students during tests, reports, lab work, and questions - this behavior seems quite uncommon (only 5 incidents) but again is rather bitterly resented as:

> "When I was giving a report the Prof disagreed with what I said and he started swearing at me and calling me stupid. So far this quarter he has done this to every person who has given an oral report."
> "He passed out the first quiz, let everyone work for 8-10 minutes, then he walked around the room looking over different people's shoulders saying, Well that isn't right!, That isn't expressed correctly!, and, Well this isn't right at all!, He did this to over half the class, and then stood over them to see what they put down. This action completely disrupted the entire quiz and the grades were bad. He did this on every quiz and the final grades were the worst I have ever seen. I got a D."

In view of the above incidents, it is fortunate that this behavior seems rare. Obviously, such behavior on the part of the teacher could only have effects on student achievement. In addition, it is probably that students would dread future quizzes and probably generalize their attitudes to the entire course and the professor. Such behavior is not in need of experimental test; the studies already reported of behavior under stress are sufficient to show the serious adverse effects of such continued harrassment.

(6) Demoralizes students by threatening punitive action - this behavior refers to both class work and personal behaviors by students. In the case of the latter it is enforcing conformity as:

> "Made me cut my hair before I could take the final."

and with the former:

> "Told the class on the first day that 1/2 the class would fail and the other half would get C's and D's if we didn't behave."
> "He made his tests so hard that few people got more than half the points, 20 of 26 people flunked."

This behavior seems to be fairly common (4% of the Category incidents) and the remarks for sub-category (5) above apply here. The constant threat hanging above students appears to be demoralizing and would result in performance decrements. Further, the intrusion into personal preferences is completely unwarranted and probably results in derogatory attitudes by other class members as well as the individual involved.

(7) Does not accept legitimate excuses or explanations - the behavior reported here often is in actual violation of school policy, that there are "legal" reasons for missing classes or quizzes and the requirement is to recognize this by acceptable actions. Some incidents were:

> "Was in the infirmary, missed the quiz. Prof wouldn't let me make it up."
> "He wouldn't accept a doctor's excuse for missing a test. Gave me a '0' on the test."
> "Had personal problems at home which forced me to turn in an assignment one hour late and he dropped me a letter grade."

This is a fairly common behavior (5% of the Category incidents) and obviously has a direct effect on a student's grade. In addition, setting such an example to students, refusal to comply with school rules, can only have undesirable effects on student attitudes. Such behavior could be corrected by appealing to administration officials but one student reported he was "afraid" to take any such action. In general, such behavior, apart from the behavioral effects on students, seems totally inexcusable.

(8) Does not know (or attempt to) students on a personal basis (e. g., by name) - this is the obverse of the large sub-category (I-1) reported under effective behaviors but seems a deliberate attempt not to know students on a personal basis as:

> "The Prof showed no respect for the students.
> He made no attempt to call a student by name.
> He would say, Hey you."
> "Prof took no personal interest in his students.
> I had him for classes two consecutive quarters,
> with others, and he never knew any of our names."

Here again is the plea for individual recognition and the "respect" that it implies. The deliberate attempt not to do so seems rare (only 5 incidents) and probably has little effect on performance but is likely serious effects on interest in and attitudes toward a given professor and course.

(9) Hesitates or refuses to answer questions (inadequate answer) - the behavior referred to is that occuring in class and is the largest sub-category here, 24.8% of the incidents:

> "He wouldn't answer a single question. Always
> said it was in the book."
> "Were thrown out of class if you asked a
> question."
> "Refused to answer questions, said they were
> a waste of time."
> "Says he doesn't know so it's not important."
> "Says will answer later but never does."

The comments to sub-category (3) above apply here. Probably many faculty members feel it incumbent to answer any question, particularly those dealing with the course material. However, many professors apparently flatly refuse to answer questions. There is no indication in the reported incidents of inability to answer questions but it is suspected that this may be a major reason for the refusals. The behavioral effects could be quite serious.

(10) Hesitates or refuses to help students (class or office) - the behaviors here refer largely to helping students with regard to class work, although help with personal problems could be included also. This is a major sub-category, almost 20% of the reported incidents. Some incidents were:

> "If a question was raised by a student he would tell them to see him after class but he always left at the end of class."
> "He said his time was too valuable to waste it on students other than during scheduled class."
> "Said, 'I have better things to do than try to explain something you should have gotten in class. Now get out of here'."
> "Said didn't have time to answer dumb questions since he had so much time involved in research work."
> "Student was sick and Prof said he would help him and when the student came to his office for help, Prof denied everything."

The hypothesis for the behaviors described here would be of major effects on student achievement and attitudes. It is a certainty in virtually any class that some students will have difficulty with portions of the course material. The refusal to give help to an individual student no matter how "dumb", is a failure to consider this difficulty plus a serious refusal to accept a major responsibility of teaching-making every reasonable effort to ensure student learning. A probable teacher correlate here is a real lack of the "empathy" for students reported in the research previously described and probably stemming from some more-or-less serious personality defect.

(11) Dogmatic and inflexible (belittles students in general) - the behavior here is usually exhibited in the classroom situation. The meaning is a refusal or inability to see any other viewpoint and, using personal denigration of students as a way of argument such as:

> "Cuss you out right in front of the whole class for disagreeing on even a minor point in what he said."
> "Professor continually informs class that they are worthless and don't deserve the teacher they have."
> "When asked why a certain method should

> be used the Prof said, Don't question what
> I say."

The behavior described here appear to be those of the typical "authoritarian personality." Probably this would have relatively minor effects on student achievement but would stifle any interest in the course and material as well as arousing fear and contempt in students. The implication is a feeling of superiority that enables one to teach students in contrast to helping students learn. Probably the most interesting feature of any research conducted with this behavior would be the personality characteristic of teachers behaving in the ways described above.

The total Category probably describes teachers who are victims of some serious personal inadequacies. In consequence, they show fear and distrust of students or, in the extreme case, outspoken contempt for them. Unquestionably student behavior can be quite irritating or frustrating at times but the behaviors described above appear to carry the implication that this is true of all students all the time. Too, there is the lack of "empathy", seemingly a real inability to appreciate the student world and its problems. One might even go so far as to infer not only a lack of concern but an actual overtly hostile attitude toward students.

It does not seem difficult to see how such teachers could show very bad effects on students; in particular, the creation of a "stifling" atmosphere could be considered quite serious from many points of view. This seems to be an important and provocative area of research on teaching: first, the location of teachers showing these "stifling" behaviors; second, showing their effects on students and, third, determining the personal correlates of these behaviors. With the latter, a further step might be to determine whether or not such persons could be "re-educated" to improve their teaching performance.

Occurence of the above behaviors seems to offer a major reason for the evaluation of teaching performance by students. They are the only persons apt to see such behaviors and, it would appear, there would be fairly substantial agreement that such behaviors should be corrected if verified.

The second ineffective behavior Category is under, "Classroom Administration" and, to a large degree, the reported behaviors are the reverse of the same category under effective behaviors.

However, there are some unique exceptions.

(1) Meets class irregularly or not at all (special sessions). Leaves lab - this sub-category is composed of fairly obvious behaviors as:

> "Missed about 1/3 of the class sessions and never told the students ahead of time."
> "Called off 40% of his classes."
> "Misses about 1/2 his classes and makes them meet at night."
> "Prof scheduled a make-up final exam and didn't show."
> "Left lab and didn't come back until the end. We couldn't ask questions."

This is the largest sub-category, 29% of the incidents, the indication being that the behaviors are fairly common. These data need no comment as to the possible adverse effects.

(2) Frequently comes to class late - this is also a quite common behavior, 24% of the reported incidents in the Category. The behavior is again obvious:

> "This Prof was always late for class."
> "Always came to class at least 15 minutes late."

This behavior is also related to (4) below, professors who are habitually late for class also tend to keep students overtime. The effects of this behavior on student achievement probably are minimal but the student irritation is not. This lack of consideration for students probably results in poor attitudes toward both the professor and course and certainly seems a needless frustration. Almost everyone will be late for class on occasion but for every class seems uncalled for; no "absent-minded professor" is that absent-minded. The general hypothesis would be of a rather slip-shod approach to the course; in fact, some students added to their incidents comments as, "then he didn't teach anything." This would appear to be an objective behavior that could be a key behavior in evaluating teacher effectiveness, speculative at present but a possible performance criterion.

(3) Permits classroom disturbances, lack of attention - this is a very small sub-category (4 incidents) and does not seem to be a generalized behavior as do the two sub-categories above:

> "People read newspapers while he lectures."
> "Prof just went on talking (in a large lecture room) while practically all the students were talking to each other."

Very likely there are few professors who would allow such student behaviors even minimally, much less to the extent reported above. However, the adverse effects on student performance, if allowed, seem quite apparent.

(4) Consistently lectures overtime - as mentioned above this behavior is related in many cases to (2) above. That is, the professor is late for class and goes on lecturing past the end of the class period:

> "Was continually late for class. Came in about 20 minutes after the hour and then kept us overtime to finish his lecture."
> "Always ran about 15 minutes over the hour."
> "Kept after the bell every day."
> "Kept a three hour class 1/2 hour overtime every day."

The comments under sub-category (2), above, apply here.

(5) Fails to state objectives and over all purpose of the course - a very small sub-category comprised of only three incidents. The incidents are fairly apparent; a failure to describe what was expected of students. Obviously, such failure can have important effects on student learning - they do not know what it is they should learn. In terms of experimental testing it is possible that such behavior could seriously affect student achievement.

(6) Makes false statements regarding course requirements and what is expected of students - in this sub-category a variety of incidents occur dealing with statements to both individuals and the class as a group. The behavior seems to be rather common, 17% of the category incidents. In many cases the professor "forgot" and, although it is difficult to believe, many of the behaviors seem to be intentional:

> "Said at the beginning of the quarter that number of cuts had no effect on our grade and then counted off for cuts."
> "Prof told class he would drop the lowest quiz. He didn't."

> "Said if I got a C on the final I would get a C in the course. I got a C but a D in the course."
> "Told me if I finished a project by the deadline I would get a B. I finished on time and got a C. Prof denied he told me this."

Such actual behaviors on the part of faculty members seems to be completely indefensible in that student performance can be seriously affected. Misunderstandings of one sort or another inevitably will occur between faculty and students but the behaviors here, in most cases, seem to be deliberate distortions of intentions. Such behaviors can only have adverse effects on all facets of student performance both in the immediate and long-term aspects.

(7) Monopolize student time with excessive or irrelevant assignments - this is a fairly common student observation, some 14% of the category incidents. The judgement is to some degree subjective on the part of the student but some of the incidents offer objective evidence that students can evaluate the behavior acceptably:

> "Handed out 50 homework problems a week which were graded. Had a hard time keeping up with other courses because of it."
> "He assigned 22 cases to be abstracted four days before the final. Our exam came from the 22 cases."
> "Handed out a quarter assignment sheet which required the students to read at least three chapters per week in addition to several outside articles. The Prof seemed to assume his was the only class the students had."
> "Prof gave large amounts of extra reading and two term papers which only counted if you didn't do them."

In the reported incidents students remarked how one course often monopolized their time with the consequence that other courses had to suffer, an actual statement of effects on student achievement. Such teacher behaviors amount to putting students in a stress situation with the possible effects discussed previously. While there is some element of value judgement in student reports of this behavior, it would seem incumbent upon any teacher to make reasonable assignments.

(8) Gives no exam before drop date - the behavior reported here is rare (5 incidents) and the incidents have obvious content along with effects on students.

The behaviors reported in this Category center around the "business-like" classroom administration that were discussed in the reported research findings and in the same category under effective behaviors. They seem to reflect a lack of pre-planning of the course to be taught along with a somewhat indifferent attitude to both the course and the students. The tone of the reported incidents indicates that students are quite resentful of the reported behaviors, in particular, where they are misled or, as some said, "lied to." In addition, these incidents comprise about 10% of the total ineffective incidents reported, an indication that slip-shod, indifferent classroom administration is by no means uncommon. It is possible that a planned program evaluating teaching effectiveness could have a major salutory effect here; often people behaving in the ways described are not aware that they are doing so until it is brought to their attention nor do they sometimes realize the real impact of their behavior. A regular reporting of such behaviors occuring repeatedly could be a large step toward improving the teaching by a given individual.

The third Category reported here is, "Student Participation" as for the effective behaviors. As will be seen the behaviors here are largely verbal discouragements to students in their attempts to "participate" in either class policies or course conduct. The behaviors are not common, 2.7% of the total ineffective incidents.

(1) Does not permit class discussion of scheduling quizzes or assignment due dates - here only two incidents were reported. Both concerned arbitrary assignment of quizzes and refusal to discuss them with the class. Apparently the usual behavior is to set an agreeable date when changes are necessary and only rarely is an arbitrary assignment made.

(2) Does not permit class discussion of material or opinion - this behavior represented virtually all the incidents reported in this Category and centers around professors lecturing and never asking or allowing for questions or actually refusing to allow questions or discussion:

> "For the entire quarter he just lectured.
> Never asked nor answered a question.
> Like a recording."

> "The worst Professor I ever had would come into class, lecture, answer questions he picked himself, and leave."
> "Student would try to argue a valid point and Prof would tell him to shut-up."
> "A student corrected a Prof one day and he threw him out of class. This curtailed student participation."

The effect of the above behaviors on student achievement is probably not major but in terms of student interest and attitudes undoubtedly has severe effects. Fortunately, the behavior seems to be rather rare, in particular, the overt discouragement of student participation. As with the authoritarianism previously noted, probably the most interesting research here would be concerned with the personality correlates of the professor who actively discourages student participation.

The fourth ineffective behavior Category, "Classroom Presence" is one where the most noticeable differences occur in ineffective versus effective behaviors. They are not, in general, the reverse of each other but seem to separate sets of behaviors. This is the second largest category of ineffective behaviors, 24.6% of the total. Most of the behaviors reported are not critical incidents in the technical sense but are in the nature of habitual behaviors:

(1) Objectional dress, manners, and appearance - here were reported a very heterogeneous set of behaviors mostly of an "irritating" nature as reported in the previously cited article by Moore:

> "He came to class drunk several times during the quarter."
> "Chewed gum while lecturing."
> "He wore the same suit all the time and the same tie, which was filthy."

The behaviors reported probably have little, if any, effect on student achievement but almost certainly affect student respect for the faculty member. In actuality, such behaviors show a disrespect for students which is probably reciprocated.

(2) Displays nervousness, ill-at-ease when talking (e. g., paces floor, easily flustered) - this seems to be a rather rare behavior (8 incidents) and probably, as above, is just "irritating" to

students rather than anything else. Practically all of the incidents described random "pacing" around the room. It is doubtful that such behaviors have any serious effects on students.

(3) Talks or presents material too rapidly - this is a fairly common behavior, about 6% of the Category incidents. The behavior seems habitual and is related to category I-9, not stopping for questions or any discussion. All of the incidents involved talking or writing so fast that students were unable to take notes and/or comprehend the material. Effects of this behavior on student achievement are easily understood as the material is "thrown-out", and students are unable to keep pace with the presentation are therefore at a real disadvantage. Such effects would be susceptible to experimental test using variously paced presentations; very likely some optimum range of presentation, neither too slow nor too fast, would be found.

(4) Lectures in a rambling, disorganized fashion - the behaviors reported are dependent to some degree on subjective opinion. A meaningful measure would be some 75% of the class members agreeing that the behavior did, in fact, occur. Again this behavior is not an incident in the strictest sense but seems to be habitual:

> "One minute he would be talking about the subject and the next he switched to something else. Nothing presented ever fitted together."
> "Spoke on one subject, changed to another, back to the first, etc. No one ever really knew what he was talking about."

The behavior reported was rare (8 incidents) but it can easily be imagined that such a disorganized presentation would be quite confusing. Probably students who did not attend such a class would show better achievement than those who did. Again the personal characteristics of teachers who show such behavior would be of particular interest.

(5) Speaks inaudibly or mumbles - this is a large sub-category, 16% of the Category incidents. The sub-category title is self-descriptive and, in essence, students reported they could not hear the lecture. The effects of this habitual behavior on student achievement need no comment.

(6) Lectures in a monotone - again a large sub-category, 15% of the Category incidents and, as above, the sub-category title describes the incidents. The most commonly reported student reaction to this behavior was going to sleep. Very likely this has serious effects on several aspects of student behavior mainly because of lack of concentration and interest. Again, an experimental test of effects would be easy to design.

(7) Difficulty speaking English - a rather large sub-category (8%) that needs no comment.

(8) Does not look at students during lecture - a rather small sub-category, some 4% of the reported incidents:

> "Stares at ceiling as he reads notes. Boring."
> "Always talks to board."
> "Stares out the window during lecture."
> "Never sees hands raised, is looking at the floor."
> "--stared at the ceiling while giving a lecture from typed notes. He would read a paragraph, start to speak, look at the ceiling, forget what he was talking about, and have to return to his notes to complete the sentence he was saying when he looked up."

It is difficult to assess such behaviors with regard to student effects. Such habitual behavior unquestionably irritates students and distracts them but there were no statements to indicate actual effects. Several students attributed the behavior to a lack of interest in them or the job but such an inference is completely speculative and will need research study for clarification.

(9) Reads majority of lecture from book or notes instead of just referring to them - this is by far the largest sub-category comprising 30.5% of the Category incidents. The sub-category title is a literal description of the incidents - direct reading from the text or notes with no attempt to embellish the material in any way. Any effects on student achievement would be a moot question but there is no question that students resent such presentations. Many of the incidents point out that class attendance was a waste of time since they too can read but usually these professors required class attendance. Students also inferred here a real lack of interest in contempt for them and the course but there is no evidence to support such inference. However, it does seem that such teaching behaviors

do meet minimum standards of responsibility toward students, if that.

(10) Uses profane language constantly - this is a very small category (7 incidents) with the obvious behaviors. The behavior is quite offensive to at least some students; in fact, one student dropped a course because of "constant cursing". Effects of this behavior on students probably would be quite selective; that is, some would find it offensive and probably dislike the course whereas others might actually enjoy such usage.

The fifth Category, "Organization and Presentation of Material", again presents incidents that differ considerably from the effective incidents under the same category. Often, it seems, a lack of knowledge of the course material is involved although this is rarely directly mentioned.

(1) Does not cover all of the course requirements - these incidents were mainly where the course was one of a sequence and are not technically incidents but are summaries of total behavior. Only a small number (8) was reported:

> "Professor does not'teach'course, he only works one kind of problem and no base is laid for the next course in the sequence. Outline, stated by school, is ignored."
> "Covered 1/3 of the course material in the first of a sequence of courses. I was hurting the next quarter."

The implications of the behavior described for student achievement are obvious.

(2) Wastes class time on trivial detail and/or subjects unrelated to course objectives - as noted, the behaviors here fall into two distinct behaviors but with the ultimate same result wasted time and effort. The behaviors are fairly common, some 23% of the Category. Some incidents were:

> "--all he talks about in class is sex and drinking. We are supposed to be studying (subject)."
> "Spent about 1/2 the quarter talking about ideas for his doctorate."
> "We spent about 1/2 the class each day talking about Viet Nam or something that had nothing to do with (subject)."

> "All he ever did was prove theorems; we never worked problems."
> "Prof spent an entire class on a problem not even in the course because a student asked question."

Student reports here were quite adamant. They resent the waste of time, particularly when it involves teachers' personal problems. Many also reported being tested on material not covered (see Category VI, below). Such behaviors may help to some degree in establishing "rapport" with students but from the number and tone of reported incidents this is carried so far that students feel it has serious effects on their achievement.

(3) Repeats material to the point of monotony (same lecture in different courses) - only a small number of behaviors were reported here (7) but students do remark the occurance and effects:

> "Lost all interest in the course; he said the same thing in the same words over and over."
> "Prof had a bad habit of saying the same thing four or five times in a lecture. Often it was not worth saying once."
> "Gave exactly the same lectures in two different courses."

The general tenor of students' remarks here center around the loss of interest in the course, quite understandably. There is probably a major effect on achievement also although this would need investigation. The point of using the same material for different courses probably reflects sheer lack of responsibility on the part of an individual faculty member and is so remarked by the students.

(4) Lectures above students' level of understanding - this is a small sub-category (15 incidents) and probably is more subjectivity-laden than any of the other evaluative statements determined by the reported incidents. Again, if this were to be taken as characteristic of a given professor, a criterion of the number of students reporting it from a given class would be required, perhaps 75%. Some incidents were:

> "Students did not know words Prof used routinely in lectures."
> "Professor taught over everyone's head, I disliked him and the course so much I switched majors."
> "Prof said to stop him if you didn't understand. I did (twice) and he said he couldn't explain it any more clearly."
> "No one could work the problems he did in class, students learned nothing (low grades)."
> "Picked the most involved examples to work; we couldn't do the simple ones."

A class conducted in the manner described in the incidents would undoubtedly have adverse effects on students from any point-of-view. In particular, where the difficulty is called to the attention of the professor any continuation of the behavior is inexcusable.

(5) Unable to work problems or answer questions - this is by far the largest sub-category here, 27% of the reported incidents. In contrast to refusal to answer questions, the professor is unable to do so. The inferences, or student's statements, were to the effect that this was a result of incompetence in the field of study. From the reported incidents there seems to be some factual basis for this. Some incidents were:

> "In the entire quarter he never worked a single problem on the board without making a mistake."
> "Couldn't do the assigned homework and when we asked the Prof, he couldn't do it."
> "Prof did not know subject, when asked a question he often put the wrong formula on the board."
> "Would enter class, work all the problems wrong and tell students to finish them; failed to prepare any class lecture."
> "--he couldn't work a problem so he got one of the students to do it."
> "Had to look in the book to answer questions."
> "Always worked example problems wrong."

The behaviors reported here are likely to have quite serious effects on all student behaviors. The inability, for whatever reason, to present the course material correctly in addition to hindering learning could result in wrong learning which is even more serious.

Further, the behavior seems to be more common than one would surmise and could have all sorts of adverse effects on students.

(6) Gives erroneous information - this is not a commonly reported behavior (7 incidents) possibly because students do not recognize it in some cases. This behavior is in contrast with that directly above in that questions are answered or information is given that is incorrect. Often a deliberate attempt to "cover" ignorance is implied:

> "--on cooling curve of metal, showed a negative curve in time. Just plain did not understand idea."
>
> "This Prof's lectures contradicted the text and when students questioned him, he said the text was right."

Certainly the kind of behaviors reported here must have adverse effects on both student learning and attitudes. The personal correlates of such behavior need to be determined, though this is a rather uncommon behavior. It certainly is inexcusable in the teaching situation.

(7) Does not or refuses to explain course material - this behavior may or may not be related to the two sub-categories above. In some cases students implied that lack of knowledge was the basis for such behaviors, but also there is another facet, seemingly, an assumption of learning or understanding that students did not really have:

> "Abstract presentation--did not provide any examples of what he was talking about - no one understood."
>
> "Would do problems on board without explaining basic method, cranked out numbers without telling the formulas used to arrive at solutions."
>
> "Never completed examples of material, just said the rest of the solution was obvious and moved on to another subject even though students expressed confusion about the previous problem."

This was a relatively rare behavior (9 incidents) as reported by students. The possible effects on student achievement seem apparent.

(8) Forces students to shoulder burden of gaining subject matter knowledge - this behavior is somewhat similar to that reported above, but here the attempt is deliberate, or at least students see it

as so. There is, of course, some responsibility on the part of students, to learn material and the point at which professor-student responsibilities interact is somewhat vague. However, students do make this judgement and it is fairly common (12.6% of the Category incidents). In addition, the behavior seems to be habitual:

> "--would start a problem but not finish it. He said he wanted his students to complete the work and more material could be covered that way. Nobody knew how to finish problem."
> "Prof assumed knowledge that students didn't have such as doing integrals in the first math course and students had never seen them."
> "This Prof told us several times that he had a hard time getting the material across and we would just have to do it on our own."
> "--filled the board with formulas and you would ask him something and he would say, 'that's for you to find out when you go over your notes'."
> "Said to ask one of the experts in the department about a question during class that he didn't know."

The behaviors reported here often do appear to be deliberate, possibly related to the hostility toward students previously discussed. Again, the undesirable effects seem to need no comment.

(9) Lectures do not contain any material not fully explained in the book - only one student reported this, possibly because it is covered in other words in the other sub-categories in this section.

(10) Hurries through course schedule without regard for student understanding of material - here only seven incidents were reported, and are related to Categories I-9 and I-10, that is not answering questions or giving help because there is not time to do so. As with those categories, the unwanted effects upon students seem obvious:

> "If he found himself behind he would just jump to the material where we were supposed to be and tell us to find the other material in the book."

(11) Lectures consist of copious blackboard notes - here only two incidents were reported; probably a behavior that is quite uncommon and whose effects would be moot.

(12) Stresses theory without explaining applicability - this is a fairly common behavior (15 incidents) and seems most noticeable with regard to student testing:

> "Lectured on theory in class and on quizzes expected students to answer highly specific questions."
> "-- proves theorems in class and on quizzes expects students to apply them with no knowledge of their application."

Here, again there is some fine line as to a given professor's responsibility for guiding students. The incidents cited here imply no guidance whatsoever in the use of the course material which seems to be at one pole, the other being "leading by the hand". There were no incidents effective or ineffective regarding the latter and, for evaluation purposes, probably only the extreme behavior could be used. This would appear to be a rather difficult behavior to assess in terms of effects on student performance, especially achievement, since the better students would be likely to learn in spite of such behavior. A very close control of student ability would be required in any experimental test.

(13) Unprepared for class - a rather common behavior (13 incidents) and again involving some degree of value judgement on the part of students. Some teachers could be "unprepared" for class in the literal sense and still give a meaningful presentation because of their knowledge of the subject matter. The distinction is likely to be made on some personal correlate of ability to deal with such a situation, that is, being unprepared and still presenting a worthwhile lecture:

> "Always leaves notes, book, etc. in office but tries to lecture anyhow."
> "Jumped from topic to topic and when students asked questions could not relate topics."
> "Talked about anything; never had a prepared lecture on the subject."
> "Just stood in front of the class and leafed through the book and would say there's stuff on page---that you should read."

This complete abrogation of teaching responsibilities hardly needs any comment.

The behaviors reported here seem to stem from two sources.

First, there is a lack of competence in the subject matter, just plain not knowing the material. Second, there seems to be a lack of effort to meet some minimal teaching responsibilities as not preparing lectures, rushing through the material, or ensuring that students are grasping the material. It is possible that some of these behaviors stem from a given teacher not being aware of just what is occuring, but some of the behaviors are quite deliberate, *e. g.*, referring students to the book and refusing to explain.

The distinction is vital because in the one case it would be possible to point out and correct these ineffective behaviors but, in the other, any remedial efforts are likely to be resisted. Certainly, the detection and alteration of such behaviors seems a desirable objective since the behaviors described can hardly have anything but negative effects on students. Probably the only way to attain the objective is through some systematic reporting method by students; single or sporadic student reports tend to be discounted.

"Evaluation of Student Performance" is the largest Category among the ineffective incidents, some 25%, and beyond question the most emotion laden on the part of students. Since grades are *the* criterion in evaluating undergraduate performance, this reaction is quite understandable. Indeed it is patent that grades are what guide student learning in most cases rather than any consideration of intellectual or personal development. No matter what one may think of this situation, the fact is that people tend to show behavior which is rewarded, in this case, by grades. If some other behavior on the part of students is desired, it will be necessary to develop the appropriate criteria to elicit such behavior.

(1) Tests students on material that was not assigned - this is the second largest sub-category, 11.7% of the reported incidents. Some degree of student judgement is obviously involved in assessing this behavior but the fact that such judgement can be made seems real enough in view of the number of incidents reported:

> "Would tell you *exactly* what to study for a quiz and his questions were never on this material."
> "--he allowed us to bring notes for the final but the questions were from another course."
> "--over half your grade came from how well you spelled, punctuated, and phrased or how neat and fancy your cover was but never told us he would grade on these."
> "Had things on quiz we had never seen before, he said that was too bad."

> "Missed two weeks of the course and tested us on uncovered material as soon as he came back."
> "He gave out a list of eight topics which he recommended very strongly that we know. In fact, they will be your quiz. Only one question out of five on the test even remotely concerned any of the eight topics."
> "Questions on the first quiz came from a chapter he said we would not be responsible for."
> "He assigned, collected, and graded homework before we covered the material in class."

The general effects on student achievement, not to mention attitudes, are so patent as to need no comment.

(2) Tests do not include material emphasized in class (covered) and/or in reading assignments - these behaviors are different from the above in that the professor's emphasis in class or deliberate statements lead to false expectations by students as to the probable content of tests. There does not seem to be a deliberate attempt to mislead, in most cases, but the ultimate effect is the same. This is the largest sub-category, 13.7% of the incidents:

> "Prof read from books as class work. Tests did not have any of this material on them."
> "Prof told a class a quiz would come solely from outside readings; it was on class notes."
> "Prof spent four classes out of six on a topic in (subject); quiz had no questions on topic."
> "Required us to read seven papers in library. Did not ask a single question on quizzes about them."
> "Assigns 'impossible' homework problems that are the basis for quizzes and count toward the final grade but he never returns problems or discusses them in class."
> "Prof talked about whatever students wanted to rather than course material. Then gave detailed quizzes over the course material."

Behaviors such as the above can have no justification but it is of interest that they do occur and so frequently. The needed research again appears to center around the correlates of such behavior on the part of faculty members. There is a possibility of some student-professor misunderstanding, in at least some cases but there also appears to be some disorganization or lack of course planning, in particular, with relation to course objectives. Whatever the cause, there is both a serious effect on possible student achievement and a rather bitter resentment on their part.

(3) Tests either exceed the difficulty level of the material or are too easy - here nine incidents were reported and the difficulty of accurately judging this behavior recurs:

> "Open book tests and copied answers out of the book were considered correct."
> "Tested on only most difficult problems."
> "Nobody understood test. 24 of 26 people flunked."

An interesting point is that students reported "too easy" tests; there does not appear in the incidents any particular reason for these reports. The teacher behavior here possibly can only be assessed by extreme cases *e. g.*, "everyone flunks or passes." Such behaviors, as above, show achievement effects and probably seriously demoralize students where the tests are much too difficult. At either extreme, the responsibility of the professor to take corrective action is apparent.

(4) Tests require memorization rather than demonstrating knowledge of principles or ideas - this is a fairly large sub-category, 6.5% of the incidents and seems to involve student recognition of what the "real intent" of education may be. There is also some connotation of their being treated as adults by appropriate testing:

> "Prof made us memorize a table of gas pressures for a quiz which he could easily have given us and allowed us to concentrate on how to work the problems."
> "His tests were mimeographed from the student manual. We just memorized the answers."
> "Gives answers to questions in class and you must give them back to him *word-for-word* on the test or they are wrong. Like high school."

> "For example, he gave us his own views on Viet Nam. On the quiz you had to give back his views or the question was marked wrong."
> "Required us to memorize lists of facts like dates. There was no apparent relation between the lists. Treats us like kids."

In evaluating this behavior some fine, possibly impossible, distinctions might be required except in the most extreme case. A knowledge of facts is indispensable to at least some degree and measurement of the effects on students here is fraught with all sorts of difficulties, in particular, some definition of parameters. Any hypothesized effects on student behavior are much too speculative to consider with the information available.

(5) Time to complete test is inadequate - this is a small category, only six incidents, and with obvious behaviors reported. There is very likely a wide student variation in reporting such behaviors which is probably closely tied to student ability. In addition, this could be used as a psychometric device where only the more able students finish the quiz and grading is thus made more defensible.

(6) Uses same test questions every year - here the complaint is actually that some students obtain copies of a repeatedly used test and thus grades are unfair. Obviously, precautions need to be taken to avoid such a contingency, however, only three incidents were reported of this event.

(7) Refuses to (does not)discuss or explain returned assignments, tests, projects, or grades - this is a fairly commonly reported behavior, 8.7% of the Category incidents and mainly involves student complaints that in not knowing what they do wrong, they have no way of learning the correct thing:

> "Prof said we had not learned anything after a test with an average grade of 34 out of 100. The class wanted to remain for the rest of the hour and discuss the test and what we had missed. He refused, dismissed the class, and started something new the next class."
> "Never returned homework after grading so you could see what you did wrong, just the grade."

> "Prof gave five quizzes during the quarter but we never saw them, just the final grade."
> "Prof refused to discuss anything related to how he graded and stated he would automatically take off 5 points if anyone questioned his grade."
> "Had 3 1/2 questions of 4 right on a quiz. Received a D and the Prof wouldn't change it or indicate why."
> "--he wouldn't even tell you why he subtracted points on your quiz, there were no marks on the paper except your grade."
> "--he gave me the wrong grade (number of points) and then would not change it without giving any reasons why."

The tone of the incidents (along with somewhat colorful descriptive language) leaves one in no doubt that the behaviors as reported above are infuriating to students. Not only is their achievement wrongly judged but the professor arbitrarily refuses to correct some obvious injustices. There is no indication as to the motivation for this sort of behavior on the part of faculty members; however, there does not seem to be any adequate defense for it.

(8) Does not comment on returned papers (tests) - to some degree this behavior is related to that described above. The only real difference is here that professors do not refuse to give the bases for grading. Only 12 incidents were reported concerned with this behavior, but here too students resent such behavior. The reported incidents are obvious but not quite as blatant as those above. This failure to inform, rather than refusal, is merely one of degree and is equally difficult to understand. In both cases the student is put in the position of being wrong but without knowing why. Both categories are possibly related to the authoritarianism discussed previously.

(9) Returns papers late or not at all - here a fairly substantial number of incidents was reported, 6.7% of the Category. The behavior does not seem to be a deliberate attempt to evade explanation but more in the nature of failure to recognize student interest possibly because of "lack of time:"

> "Did not return quizzes or homework, just gave a final grade."
> "--gave mid-term and did not return it. Said he would bring it to next class when questioned

> but did not do so. Finally brought it in three weeks later."
> "We had two quizzes and a mid-term and never got any of them back. We didn't know how to work for improvement."
> "Prof failed to return first test before giving second. Test had some of same questions; you could miss the same question twice."
> "Gave his one test back four weeks later."

Students appreciate the possible effects of this behavior on their achievement - the inability to improve if you do not know what it is you are doing wrong. In addition, there is considerable emotional toning, incidents calling names or commenting. Such comments seem fully justified.

(10) Does not grade quizzes or assignments - only nine incidents were reported and with obvious behaviors, as giving a quiz but not collecting the papers. This behavior is difficult to explain but it is not difficult to imagine adverse effects on students, especially a lack of interest.

(11) Grades on classroom participation only - only one incident was reported in this sub-category, although such participation is apparently graded in part in many classes. An interesting point is just how this student behavior is graded.

(12) Grades on irrelevant characteristics (dress, major, sex, own biases, etc.) - if any one professor behavior is seen as indefensible by students, this is it. Even the recipients of the favorable treatments seem to resent it. Fortunately, it seems to be relatively rare, 4.4% of the Category incidents.

> "Did well on all the tests but got a C. Checked other grades and found only majors got A's."
> "Student borrowed a book from the Prof who wanted it back in 10 days. On the tenth day could not find the Prof, took the book in the next day and was told he would get an F in the course, which he did."
> "A girl in the course had lower grades than I did all quarter. She got an A, I got a C."
> "Took student to his office and said 'if you don't quit wearing those clothes, I will

> flunk you'."
> "Final consisted of 2 problems. Everyone got the first one right and missed the second. Somehow the Prof gave A's, B's, D's, and F's."
> "Gave me an A instead of a B because he said I got along with him."
> "Marked a problem wrong for having a/ax rather than $\frac{a}{ax}$."

This behavior seems to puzzle students as much as anything else. For many it seems this is their first encounter with such arbitrary behavior, at least with full awareness. It seems to evoke contempt for persons showing such behaviors but effects on achievement are not reported, possibly because most such behaviors are after the fact, i. e., final grades.

(13) Grades on class attendance - incredible as it may seem, some professors seem to give grades only on class attendance. Further, these incidents represent almost 10% of the Category where either all or a major portion of a student's grade is a result of his class attendance:

> "He allowed two cuts and then you flunked."
> "--called roll at end of quarter. If he recognized you, you got your average grade, if not, he dropped you a letter grade. He didn't recognize me."
> "He missed a class and rescheduled it. One student couldn't make it and he reduced his grade one letter."
> "He missed 1/2 the classes but graded students on attendance."
> "100% attendance required to pass."
> "Reduced your grade one letter for each cut."
> "Gave A's if you missed no classes, B's if you did."
> "Took roll 3 times, if you weren't there he assumed you had missed 1/3 of the classes and the best grade you could get was a D."

The best that can be said for this policy is that it is objective (see exceptions above). The effects on student achievement of this policy are not obvious but there is no doubt of the resentment at such arbitrary behavior; often it is in clear violation of school policy. Possibly also this is a substitute for an interesting course, some student comments make this appear to be the case.

(14) Grades on final exam only - the behavior is the single one indicated and students feel that it is unfair. Very likely they are right in that the reliability of such a test is likely to be low and, further, other work of any quality goes for nothing. There is little question of the effects on student attitudes of this behavior but, in terms of achievement, would be difficult to predict. Only 10 incidents were reported.

(15) Grades are not in accord with test scores - this is not a common behavior, 10 incidents, but is bitterly resented. Often it is after the fact and the student has little in the way of recourse. Some incidents were:

> "Prof gave me a D with a 68 average and he gave a couple of people I knew C's with 63 and 66 averages."
> "I had a 76 average and a friend 47; he got a B and I got a C. I asked the Prof about it and he said there were a wide range of C's; whatever that means."
> "I passed every test and was flunked."
> "Had a 68.8 average in class, a B in lab, and got an F. Only one person had a higher grade and nobody else flunked."

One is reluctant to believe it, but in these incidents there is the connotation that such grading is directed at a particular individual. All grading leaves something to be desired in terms of reliability but here, with quantitative evidence available, a student receives an unjust grade - there seems to be no other explanation than personal discrimination. Since this usually occurs after the fact, effects on achievement are nil but one cannot help wonder about the campus reputations of such professors. Students do know other person's grades.

(16) Does not give credit for partially correct answers - here about 3.5% of the incidents for the Category were reported. Probably all faculty members have had the experience of haggling with students about the correctness or incorrectness of answers and probably this is an important motivator of the behavior score right or wrong and no

haggling. However, students do remark the behavior and consider it capricious. The incidents reported obvious behaviors and are somewhat subjective. It would be most difficult to evaluate this behavior accurately because of this subjectivity and its effects on students equally difficult to predict. However, in terms of frequency of mention, it seems to be a teacher behavior that is in need of investigation.

(17) Passes and/or fails, or gives grades to a predetermined percentage of students in class (or large percentage) - this behavior seems to come from some arbitrary standards (seen as absolutes) by some professors. Even before the class starts, a grade distribution is known. It is a fairly common behavior, 6.7% of the reported incidents:

> "Never gives a grade above a C."
> "Told class on the first day that he never gives A's or B's."
> "Final grades were one C, two D's, and 12 F's."
> "--discouraged good work by giving all students C's on homework."
> "Student had a B average during the quarter, B on final but got a C in the course. Prof made up an artificial distribution since he was too lenient and everyone was getting A's and B's."
> "He always flunks 1/2 the class."
> "--40 of 50 people flunked the course."
> "Prof said at first lecture, I am going to flunk 50% of the class and will give no A's and B's."
> "--a proper distribution requires me to flunk 20% of the class."
> "Gave everyone a C."

Restrained, unemotional comment here is rather difficult. Certainly the effects of such behaviors on students must be most demoralizing and discouraging. They recognize the complete arbitrariness of such behavior (but not the ignorance behind it) and the incidents have a note of almost despair. It would seem incumbent upon any administration to stop such behavior on the part of faculty members, even academic freedom is not an absolute.

(18) Does not curve grades - here the implication is similar to that above, *e.g.*, arbitrary grades but there is not the tone of

vindictiveness or hostility exhibited. The professor seems to accept his evaluation of student performance as infallible. However, students do not agree:

> "Grades whole class on a 'national distribution'(?) with 4%-A, 10%-B, 40%-C, 30%-D, and 16%-F."
> "Prof would not curve grades even though only one student passed test."

Other incidents were similar to the latter, in some cases entire classes failed. Again the extreme effects on student behaviors seem apparent.

(19) Does not check accuracy of student grader - this incident was reported only three times and had to do with a student seeing his grade was wrong. Probably some such mistakes are unavoidable even where an attempt is made to check scoring.

(20) Make-up tests are made excessively difficult - only three incidents were reported here. All were subjectively stated so that it would be difficult to determine exactly what did occur; however, such a possibility does exist in view of some of the behaviors reported above.

(21) Gave no quizzes and/or final - here again only three incidents but it is a puzzle how meaningful grades can be assigned in such a course. Effects on students cannot even be imagined.

(22) Refuses to change an incorrect grade - this behavior refers to incidents where the student clearly did not deserve the grade given, 7 incidents:

> "--accused student of copying material. Showed my library notes and references but would not change F grade."
> "--showed from book answer was right but he would not change it."
> "--showed he had misgraded. He dropped me a letter grade."
> "Prof called me a liar when I said I had turned reports in. He later found them but still gave me an F."

These reports, fortunately, are isolated but their effects on individuals can be serious or even traumatic. The students reporting them did so in quite emotional tones and probably no one would defend such behavior. Again this sort of behavior would seem to spring from some basic personality defect on the part of the teacher.

The last sub-category of miscellaneous incidents involved evaluation of student performance but were not classifiable under any of the above, as:

> "--my classmates turned in the Prof's mimeographed solutions to the homework assignment with my name on them. The Prof handed them back the next day and wrote on my paper, 'Good work but it lacks originality'."

Hardly any more comment is needed on the quality of evaluation of student performance, especially in view of the preceding sub-categories.

The last major Category under the ineffective behaviors is unique to the set, "Interest In the Job of Teaching". It is a small category - less than 2% of the total ineffective incidents - and, it is to be hoped, the behaviors are as rare as this might indicate:

(1) Makes derogatory comments about teaching - this is behavior that indicates a real dislike for teaching apart from any other interest or consideration:

> "The Prof told us he hated his job and didn't care what we did or learned. He said we didn't have to come to class much but we'd probably all get A's and B's."

Here only three incidents were reported, with little in the way of student evaluation. The effects on student performance and attitudes are obscure.

(2) Belittles value of the course he is teaching - these represent about 1/3 of the total Category incidents and largely seem to reflect a person teaching a course at department direction:

> "He told the class he didn't like what he was teaching, and the class would just have to read and learn the material themselves."

> "Told students the course was useless."
> "Told the class he couldn't care less about the course and the students. He lived up to his word."

Hopefully, it will never be known what effect this behavior may have on student behavior because no one would want to experiment with it as a major independent variable and it is so uncommon that it would not appear regularly enough to be evaluated. However, it probably has quite serious effects on both achievement and attitudes where it does occur.

(3) Criticizes fellow teachers - only two incidents were reported in this sub-category and were rather subjective. Probably most of this behavior, where it does occur, is not done publicly. In general, it would seem to have little or no effect on students.

(4) Primary interest is consulting or research - this behavior is taken from direct statements by professors where they indicated to students that classroom teaching was detracting from their real interest. The statements can easily be imagined as can effects on students. Actually such behaviors are a direct insult to students and, in a sense, a confession of a somewhat distorted approach to their own life adjustment. Probably most faculty members have more-or-less lengthy periods where classroom teaching has seemed onerous but here we have the admission of "just doing it for the money" instead of working in field of interest.

This behavior may have little or no effect on student achievement but does evoke quite vigorous resentment and a recognition of the fact of being "used". Again this seems to be stretching "academic freedom" beyond any reasonable point.

The behaviors reported under the ineffective incidents, in general, could hardly have anything but undesireable effects on student achievement or attitudes or both. Probably one of the more undesirable effects would be the generalization of such attitudes to form a stereotype of college faculties. As far as the author is aware, there is no evidence to support this but the behaviors reported above seem sufficient to warrant comprehensive investigation of just what is occuring in college classrooms.

Many faculty members tend to dismiss student complaints as somewhat exaggerated, coming from a few disgruntled persons, or even as purely imaginary. From the data here it would seem that there is a broader basis for student complaint than has been assumed; ineffective behavior reports actually exceeded effective.

Some student complaints or protests are undoubtedly a result of personal views or idiosyncracies but, for the most part, those reported here are factual events that actually occured. In addition, there are a great many of them covering a wide range of behavior. It can be submitted that here is sufficient evidence that a planned program of faculty evaluation is a basic requirement if students are to receive adequate classroom teaching.

DIMENSIONS OF EFFECTIVE AND INEFFECTIVE TEACHING PERFORMANCE

Some Statistical Considerations

Before the attempted integration of the effective and ineffective behaviors reported, it is well to show the representativeness of the incidents and any biases in the data.

Analyses of the categorization system were made by sex of the student reporting the incident and by class standing. The results are shown in tabular form below.

Table 2

Reported Effective Incidents by Sex

	Category						
Sex	I	II	III	IV	V	VI	Totals
Male	344 (29.0%)	72 (06.0%)	113 (09.5%)	56 (04.7%)	387 (32.8%)	210 (17.7%)	1182
Female	45 (40.9%)	8 (07.2%)	11 (10.0%)	9 (08.1%)	25 (27.6%)	12 (10.9%)	110
Totals	389	80	124	65	412	222	1292

The above distribution was subjected to a chi-square test and a significant difference was found between men and women in reporting categories of effective teaching behavior. (chi-square was 13.40 with 6 degrees of freedom, significant at the .02 level).

The women reported more incidents in Category I, Personal Relationships With Students, than would be expected by their distribution in the interviewee population. They were also somewhat higher in Categories IV, V and VI but the major influence was Category I. This finding is also characteristic of women in job satisfaction studies; they tend to emphasize personal relationships with supervisors much more than men. This finding is, however, in contrast to ratings of teacher performance where sex differences in such ratings have been minimal.

At present there is no adequate explanation for this finding, either here or in the job satisfaction studies. Possibly it is a reflection of past learning of men and women in coping with their environment.

Table 3

Reported Effective Incidents by Class Standing*

	Category						
Class	I	II	III	IV	V	VI	
Freshman	58	10	21	9	46	33	177 (15.0%)
Sophomore	70	21	18	12	77	32	230 (19.5%)
Junior	79	19	32	17	108	59	314 (26.6%)
Senior	128	25	38	20	143	72	426 (36.0%)
Graduate	8	2	4	3	13	4	34 (0.2%)
Totals	343	77	113	61	387	200	1181

*Incidents from students admitted as high school students in a special program at Georgia Tech have been omitted.

A chi-square test of the Table 3 distribution showed no differences in reporting incidents by class standing. Freshmen showed a somewhat discrepant ratio, high in Category V, and Sophomores in Category II but, in general the discrepancy ratios were quite low indicating good agreement among individuals from different classes.

Table 4

Reported Ineffective Incidents by Sex

	Category							
Sex	I	II	III	IV	V	VI	VII	Totals
Male	237 (18.6%)	140 (10.9%)	36 (02.8%)	321 (25.0%)	207 (16.1%)	321 (25.0%)	18 (01.4%)	1280
Female	27 (29.6%)	9 (09.8%)	1 (01.0%)	17 (18.6%)	15 (16.4%)	21 (23.1%)	1 (01.0%)	91
Totals	264	149	37	338	222	342	19	1371

A chi-square test of the Table 4 distribution indicated no significant differences between men and women in reporting incidents in the various Categories (chi-square = 8.13 with 6 degrees of freedom). Women were again somewhat high in Category I as above, but not significantly so. As can be seen, Categories I, IV, V, and VI accounted for the bulk of the ineffective incidents.

Table 5

Reported Ineffective Incidents by Standing

Class	Category I	II	III	IV	V	VI	VII	Totals
Freshman	27	12	6	37	39	50	0	171 (13.4%)
Sophomore	52	31	7	65	38	64	2	259 (20.3%)
Junior	66	48	7	89	50	86	4	350 (27.5%)
Senior	89	50	10	118	62	123	11	463 (36.3%)
Graduate	4	3	2	10	10	4	0	33 (02.7%)
Totals	238	144	32	319	199	327	17	1276

A chi-square test of the Table 5 distribution is not strictly permissible because of two zero cells. However, with such large numbers, and in lieu of collapsing Categories VI and VII, it was calculated and found to be insignificant (chi-square = 26.85 with 24 degrees of freedom). The only major discrepancy was Freshmen were somewhat lower in reporting incidents in Category II, otherwise, the observed-expected discrepancies were quite small.

The above results indicate some possible bias in reporting behavioral incidents according to the sex of the respondent. Women tend to emphasize the personal relationships of student with instructor, in particular, with regard to the effective behaviors. At this point it is unknown whether or not this bias will carry over to actual evaluation of teacher performance. The question is would women remark and record those behaviors concerned with interpersonal relationships while failing to note the behaviors reported in the other categories. Such a determination would require an experimental test of the developed evaluation instrument in a mixed class of men and women. A possible biasing consideration is present in the contextual situation of data collection. There are relatively few women students enrolled at Georgia Tech, in the period under consideration they were less than 2% of the student body. As noted under the ineffective incidents, they are subject to some discriminatory treatment and, in consequence, may be more prone to note the student-professor relationship. Again empirical test of this possibility is required.

Class standing had no effect on reporting either effective or ineffective incidents. In effect, the same behaviors are seen by students at all levels with the inference that the behaviors observed can be remarked and evaluated by all students. Another point here is that the student sample was heavily biased with regard to upper classmen. This bias was intentional with the thought that such students have had more opportunity to observe a variety of teaching behaviors. The fact that class membership bias did not in turn bias reported incidents is encouraging to the acceptance of the behaviors reported as being consistently seen in a variety of classroom situations.

It had been planned to do an analysis similar to those above for the majors of persons reporting incidents, along with one for age. In the case of the latter the age distribution was so narrow that no possible effect could have been shown. With the former, there were so many majors reported that no such analysis was possible. Data describing the reports, by major, are presented in Appendices 2 and 3. The rank order correlation (rho) of numbers of effective and ineffective incidents by the different majors was found to be 0.95.

In general, it appears that a closely representative sample of behaviors has been collected with the reported incidents. The sex bias found is from only 45 incidents of the total reported and, while it could be a real phenomenon, may be a function of the contextual situation of incident collection. If later empirical tests should verify the finding, any performance evaluation would need correction for the bias. With this consideration in mind, an attempt is made below to integrate the effective and ineffective incidents into a rationale as a basis for a teaching performance evaluation device.

Personal Relationships With Students

This general Category of teacher effectiveness has received full recognition in the rating, factor analytic, and critical incident studies discussed in the literature review. It has been remarked as "ability to mix with students", "relations with students" and "Empathy". The variety of behaviors reported here make it difficult to conceive of any single term to cover them or any single teacher trait as their basis. Measured authoritarianism has been shown to distort both teacher perceptions and behaviors by Gowin and Payne (1962) op. cit., but it hardly seems possible that it would account for all, or even a major portion, of the behaviors shown. A summary of the behavioral statements is shown in Table 6.

Table 6

Behaviors--Personal Relationships With Students

1. Knows students by name; deliberate attempts not to know.
2. Shows no favoritism toward certain students or groups.
3. Gives help; refuses to do so. Holds special "help" sessions.
4. Answers questions (refuses to do so); ridicules.
5. "Rides" certain students.
6. Mingles with class; discusses extra-class issues.
7. Loses emotional control.
8. Harasses students.
9. Punitive threats.
10. Dogmatic; inflexible.

For some of the above there are both positive and negative behaviors but for others only the one aspect. However, the behavioral set appears to be much broader than "empathy" or some generalized benign feeling toward students. There is the behavior of giving class and individual help, often at considerable personal sacrifice. There is the loss of self-control and deliberate harassment of students. It is doubtful if any professors would show all of these behaviors, positive or negative, however, when the behaviors are shown they seem to be the habitual with a given individual.

On the positive side might be hypothesized a stable person deeply committed to ensuring student learning, with a real appreciation of individual differences in learning a given subject, and an interest in students as persons. The negative of this would be hypothesized as having several possible sources, e.g., lack of emotional stability, some personal inadequacies, authoritarianism, or possibly some serious personality defect in the extreme case -- deliberate ridicule or harassment and threats to students.

In general, the Category is important to students (some 30% of the effective and 19% of the ineffective incidents). To experimentally assess the behavior it is probable that teacher characteristics (personality traits) would need to be measured and related to classroom behavior. The negative behaviors do not seem to be the result of ignorance of teaching practices nor of not being aware of the impact of one's behavior, for example, outright refusal to answer student questions. Such behavior, along with many of the other negative behaviors described, can only be deemed deliberate. Such deliberate acts probably spring from some personal incentive.

Classroom Administration

This dimension of faculty member performance has also been repeatedly remarked in previous studies, both the positive and negative aspects, particularly in the factor analytic studies. It accounts for slightly over 6% of the effective incidents and 10% of the ineffective incidents. The difference is likely accounted for by the ineffective being more noticeable than the effective behaviors. Some of the major behaviors are presented in Table 7.

Table 7

Behaviors--Classroom Administration

1. Attends (on time for) all classes; misses (late for) many classes.
2. Always ends class on time; always runs over time.
3. Presents study plan; no plan or false statements.
4. Grades, grades promptly; does not grade.
5. Gives examples of quiz questions.
6. Gives excessive work; permits classroom disturbances.

The basic teacher correlate here seems to be a preplanned class before the class actually starts. When combined with the class attendance by the professor, the general trait might be conceived as some sort of interest in the course or the obligation felt toward students and the job. However, the negative aspect of monopolizing student time does not seem congruent with these. It would be hypothesized that professors here fall into at least three groups, the positive and two negative groups. A part of this behavior probably could be assessed objectively by the existence versus non-existence of a study plan for the particular course. In addition, class attendance and tardiness are objectively measurable or reportable.

In general, this is one of the more easily evaluated faculty performance dimensions and should be one of the more easily correctable. Simple insistence to appear on time for all classes by department administrators would remove much of the negative behavior. The complaint of a professor requiring excessive work for a given course is another matter. Some of the incidents reported leave no doubt that this behavior does occur, however, in many cases there would be a large element of subjective judgment in evaluation of the behavior. Possibly this would reduce to evaluation of or, alternatively, having evaluators indicate the basis for opinions of excessive course work. Whatever the evaluation mechanics, it would appear this is an area to be evaluated because of its importance even though, relatively, it is a rather infrequently encountered behavior.

The third behavioral Category concerning class participation in either policies or subject matter discussions, or both, showed 9.6% of the effective incidents and 3% of the ineffective. In the effective case the professor asks for democratic decision as to quiz dates, for example, encourages class opinion and discussion of the subject matter, or criticisms from students as to his teaching effectiveness. The ineffective incidents concerned not allowing or active discouragement of such student participation. This latter is an uncommon behavior on the part of faculty members and the former seems to be the characteristic manner of conducting a class on the part of a limited number of faculty members.

There is some evidence, as reported in the prior literature survey, that this method of teaching is effective when used by certain teachers but not when used by others. Students seem to appreciate its usage but such usage is probably limited by the nature of the subject matter, the amount of material to be covered, and traits of the particular teacher. The interaction of all these parameters with regard to their effects on student attitudes and achievement would need to be determined before it would be possible to recommend the usage of "student participation" in a given situation. At present such interactional evidence is lacking.

The fourth Category, Classroom Presence, is largely reported in negative terms, 24.6% of the ineffective incidents and only 5% of the effective. Further, the incidents tend not to be the obverse of each other. The behaviors are summarized in Table 8.

Table 8

Behaviors--Classroom Presence

Effective	Ineffective
1. Dramatizes lecture material.	1. Irritating personal habits.
2. Uses humor in lectures.	2. Tense, ill-at-ease.
3. Does not talk-down to students.	3. Presents material too rapidly.
4. Pleasing personal appearance.	4. Poor personal appearance.
	5. Rambles, mumbles, talks in monotone (reads lecture).
	6. Does not look at students.
	7. Profane

To some degree this behavior has been remarked in previous research and generally reported on a "restriction" of behavior continuum. The meaning seems to be open, good-humored behavior as opposed to withdrawn and limited behavior before a class. Possibly has some relation to emotional stability and self-confidence. The larger segment of this Category is concerned with public-speaking techniques. The rambling, mumbled monotone or "just reading the book" are quite irritating to students in addition to the fact that literally they cannot hear the lecture in some cases. This commonly observed behavior should be correctable to some degree, first, by making the teacher aware that it is occuring and, second, corrective work in public-speaking would alleviate some of the more obvious behaviors. However, the entire Category seems to be based upon some characteristic personal traits of a tendency toward open, somewhat dramatic speaking ability versus the withdrawn, dry presentation. Whether such behavior as the latter could ever be fully ameliorated is questionable, some improvement is all that could realistically expected.

The fifth behavioral Category concerned with presentation of subject matter is the most often reported effective behavior, 34.2% of those incidents, and an important one for ineffective behaviors, 16.3% of those incidents. The importance of this behavior to student learning seems obvious and, in addition, students seem to be able to recognize its occurance and non-occurance. The behaviors reported in this Category are specific to the material in some cases but a general theme seems to underlie all of them -- a conscientious attempt to ensure that students learn as well as possible. The behaviors are quite heterogeneous and, in some cases, only the effective or ineffective behaviors, with no obverse were reported. A tabulation of such behaviors is shown in Table 9.

Table 9

Behaviors--Organization and Presentation of Material

1. Works problems, home-work, etc.; cannot (refuses to work problems.
2. Stresses important points; covers trivial (wastes time).
3. Presents examples (current, personal); lectures on theory only.
4. Gives hand-outs, uses visual aids, field trips, outside lecturers.
5. Asks class questions; admits not knowing answers.
6. Repeats monotinously.
7. Does not cover course requirements.
8. Students shoulder burden of learning.
9. Lectures over students' heads.
10. Gives erroneous information.
11. Not prepared for class.

The fact of a large number of effective incidents reported in this Category is an indication that students can recognize these behaviors when they occur and, in addition, the importance to students of this teaching effort. Basically, the behaviors seem to stem from three sources, first, is just sheer preparation for the class and understanding where students might have difficulty. The second centers around competence in the subject matter, answering student questions seems the best indicator here. The third is an effort on the part of the professor to determine if students do understand the material (asking questions of students or asking them for questions).

No professor is likely to show all these behaviors either positive or negative. For example, field trips would be irrelevant in some subjects. The behaviors reported seem to be the result of varied personal characteristics. One is a conscientious effort to present material as clearly as possible and an attempt to determine student understanding. A second is cognitive in judging the relative importance of material and, finally, a recognition of personal and student difficulties in presenting and understanding course material. In terms of student achievement this category would seem to be the one most worthy of research, that is, just what is required of a professor to maximize student learning in a specific subject matter. The research model would be complex, a matrix of subject matter characteristics and points of difficulty, effort and understanding on the part of the professor, and student characteristics. However, this Category appears to be basic in teaching effectiveness and research effort is likely to be well rewarded.

The sixth behavioral Category involving the evaluation of classroom performance of students is surrounded by more controversy and emotion than any other. Again some of the effective behaviors do not have their ineffective counterparts. The behaviors are shown in Table 10.

The frequency of the ineffective behaviors in this Category is a clear indication of how widespread ineffective grading may be, 25% of the ineffective incidents. The bulk of these incidents seem to stem from an ignorance of some elementary psychometric considerations. One of these, tests should cover relevant course material, seems so obvious that the large number of incidents indicating many do not in fact cover relevant material is surprising. What possible justification could be advanced to explain such behavior eludes the imagination. Some of these complaints could, of course, be distorted by student perceptions but when so many are reported one can only conclude that they have some substantial basis in fact.

Table 10

Behaviors--Evaluation of Student Performance

1. Tests cover relevant material; tests irrelevant to course.
2. Tests require knowledge of principles; tests require memorization.
3. Curves class grades; does not or refuses to curve grades.
4. Will change grade if warranted; refuses to change unjust grade.
5. Arranges to improve grade (extra work, make-up).
6. Tests too difficult; too long.
7. Refuses to (does not) explain grading system.
8. Grades on irrelevant characteristics.
9. Grades on class attendance.
10. Passes/fails given or predetermined percentage of class.
11. Gives no tests but assigns grades.

The principle of curving grades is well-known to students and the related behavior of failing a predetermined percentage of a class seems to be known also. Again in psychometric terms, the establishment of some absolute standard of student performance is difficult to justify. Actually it is often a failure to recognize the complexity of student performance evaluation as stemming from course material, the professor, the tests administered, and their interactions. To establish some arbitrary standard of pass/fail in such a context is to do a grave injustice to students.

Refusal to explain a grading system, not grading student work, or failure to change a clearly wrong grade also appear to be completely unjustified. In effect, student performance is evaluated on some basis unknown to them or deliberately wrongly evaluated. Giving a grade on irrelevant characteristics, such as sex, seems so deplorable as to need no further comment.

As has been stressed throughout this report, the assessment of performance is a very complex task. In the case at hand, student performance, it is also of vital concern to those being evaluated. Here, if anywhere, it seems incumbent upon a professor to exert every effort to make his evaluations as fair and objective as possible. In fact, it appears that often this is not the case, indeed, some professors consciously give unjust grades. Two corrective measures seem warranted. First, a basic knowledge of psychometric principles is needed and, second, administrative attention to student complaints about grades. Inevitably some students will complain about their grades, tests, or evaluation of performance in general but here the consideration is to discover the factual basis of such complaints and take corrective action if grades are unjust. Unfortunately, the entire problem of student performance evaluation is often solved on highly subjective basis but even so it is difficult to question the proposition that a grade should be justifiable on some basis.

This behavioral Category also raises a question concerned with the evaluation of teacher performance. Often the student will not be aware of the relevant event until he receives his final grade, and the behavior on the part of the teacher is not observable in the classroom. The general procedure in evaluation of teacher performance has been assessment before the final grade is known. Related too, is that if students are asked to evaluate performance knowing their final grade the evaluation could be distorted by the grade received. A possible solution could be evaluation as it is usually done and after completion of the course a single Yes-No question, "Do you think the final grade you received in XXX was fair?" It is suggested that all students should be sent such a questionnaire as a routine administrative procedure, possibly a function of the Registrar. Again it probably would be necessary to evaluate by extremes as 75% of the students in a given class responding "No" to the question, in addition, the question would need careful phrasing to assess the desired performance.

In general, this Category is important for future study, in particular, its ineffective aspects. The behavior is common, crucial to students, and readily amenable to correction.

The final Category of interest had only negative incidents reported and a small number, 1.4% of the total. The behaviors reported were mainly those adversely critical of the course being taught and an expressed interest in consulting or research rather than teaching.

It is difficult to understand the motivation for the reported incidents. What could a professor hope to gain by making such statements? They are insulting to students and have no observable positive effects. Fortunately, this sort of behavior seems to be quite rare. The opposite behavior of interest in and enthusiasm for teaching was not remarked per se by students but is implicit in many of the effective behaviors described. The general inference of "interest in teaching" has been remarked in many past research efforts but, seemingly, will have to remain an inference from behavior since the positive statement by faculty members does not seem to be observed.

The above categories can be considered the dimensions of effective and ineffective teaching performance. As described, they do not all necessarily occur in the classroom and tend to be habitual behaviors rather than "incidents". The effective and ineffective incidents are in some cases polar behaviors, but not always so. Further, at least one of the behaviors, fairness of grading, cannot be meaningfully evaluated until the class is completed, including assigning a final grade. These considerations enter into the construction of an instrument to assess teacher performance and the administration of such a device in actual use. The developed evaluation questions, shown in the next section, will reflect these considerations.

EVALUATIVE QUESTIONS
EFFECTIVE AND INEFFECTIVE TEACHING PERFORMANCE

The questions developed to evaluate the reported effective and ineffective teaching behaviors are presented under their applicable categories below. In the actual use of an instrument developed from these questions, it would be better to eliminate the category headings and present the questions in a random order. The intent would be to prevent, as much as possible, biases and stereotypes from influencing student responses.

The question format is of a Yes-No type asking whether or not the behavior occured. In some cases one question incorporates both the effective and ineffective behavior, that is, the behaviors are mutually exclusive, if one occurs the other cannot occur. In other cases only single questions are presented where only the effective or ineffective behavior was reported.

It is presumed that any actual form based upon the questions below would be headed with a statement as, "Did the professor in this course:" The individual questions would then be presented with Yes-No columns at the right of the page as is illustrated by the first question. The intent is to present the student with so-called "action statements" for evaluation similar to those in Table 11 below.

Table 11

Evaluative Behavioral Statements

I. PERSONAL RELATIONSHIPS WITH STUDENTS

1. Know or attempt to know students' names? Yes___ No___
2. Talk with students before and/or after class?
3. Hold social events for his students?
4. Give advice or assistance at student request (class or office) with personal problems?
5. Discuss (answer questions on) extra-class issues?
6. Compliment students on good answer?
7. Encourage (answer) all questions in class?
8. Treat all students equally (regardless of sex, major, etc.)?
9. Ridicule, "ride", or otherwise embarrass students (either on questions or their performance)?
10. Give or encourage individual help with course material (class or office)?

11. Lose control of himself in class (shout, curse, anger, etc.)?
12. Bother (harass) students during recitation, quizzes, etc. ?
13. Make threats concerning class work or personal behavior?
14. Accept legitimate excuses, explanations (as for missing quiz)?
15. Refuse to listen to or recognize other viewpoints in class?
16. Say or indicate in some way that students are inferior?
17. Provide special "help" sessions for course material (individual and/or class)?

II. CLASSROOM ADMINISTRATION

1. Meet all scheduled (re-scheduled) classes?
2. On time for all classes?
3. Inform class if he will be absent?
4. Discuss quiz dates or deadlines for student convenience?
5. End lectures at end of class time?
6. Distribute a course outline or study plan (course objectives)?
7. Follow course outline or study plan?
8. Give examples of quiz items?
9. Require and grade homework?
10. Return papers and quizzes promptly?
11. Permit classroom disturbances (as students talking to each other)?
12. Make false statements concerning course requirements (number of cuts, grading, etc.)?
13. Give excessive work?

III. STUDENT PARTICIPATION

1. Ask student preference as to topics covered?
2. Ask students to criticize his teaching?
3. Schedule quizzes, deadlines, etc., at the convenience of the class majority?
4. Encourage (ask for) discussion, questions, or student opinions?
5. Ask questions to determine class (individual) understanding of course material?

IV. CLASSROOM PRESENCE

1. Appear well groomed?
2. Speak clearly and distinctly:
   A. Mumble?
   B. Talk too softly?
   C. Talk in a monotone?
3. Use dramatic gestures (phrases) to emphasize important points?

4. Use humor in lecture to illustrate points?
5. Read lectures from notes or book?
6. Appear nervous, ill-at-ease during lecture?
7. Talk or present material too rapidly?
8. Give rambling, disorganized lecture?
9. Look at students during lecture?
10. Use language students understand?
11. Use profane language excessively?

V. ORGANIZATION AND PRESENTATION OF MATERIAL

1. Begin class with a review of previous work?
2. Stress, in some way, important points in the material?
3. Use current, pertinent, and/or personal examples to illustrate points?
4. Show usefulness of material in "real world"?
5. Admit not knowing answer to a question?
6. Use outside references to supplement course?
7. Distribute hand-outs/notes to supplement lecture?
8. Use visual aids to supplement lecture?
9. Provide for field trips?
10. Have guest lecturers?
11. Have full command of the subject matter?
12. Give lectures different from (supplement) text?
13. Cover all course requirements?
14. Avoid trivial detail?
15. Answer questions; work problems if requested?
16. Lecture over student's heads?
17. Give erroneous information about course material?
18. Refuse to explain material?
19. Make students learn "on-their-own"?
20. Follow course schedule?
21. Prepare for class?

VI. EVALUATION OF STUDENT PERFORMANCE

1. Base tests on relevant (covered) material?
2. Base tests on knowledge of principles rather than memorization?
3. Base test on emphasized material?
4. Make tests too easy or difficult?
5. Schedule quizzes at regular intervals?
6. Allow adequate time to complete tests?
7. Comment on (correct) returned papers, quizzes, etc.?
8. Excuse high average students from final?
9. Permit extra work to improve grade?

10. Disregard lowest test score in grading?
11. Use same tests every quarter?
12. Refuse to explain grading system?*
13. Tell how students are to be graded?
14. Curve grades either:
    A. To compare individual performance with class performance?
    B. To reduce student grades?
15. Return all papers and quizzes?
16. Grade all quizzes and assignments?
17. Give make-up tests at individual convenience?
18. Grade on such things as major, sex, athlete, etc.?*
19. Grade on class attendance?*
20. Give final grades in accord with test scores?*
21. Grade on final exam only?
22. Pass/fail a predetermined percentage of the class?
23. Try to have make-up tests excessively difficult?
24. Change a clearly unfair grade?*
25. Consider effort, participation, application in assigning final grade?
26. Use student to grade work?

VII. INTEREST IN JOB OF TEACHING

1. Make derogatory comments about teaching?
2. Make derogatory comments about the course?
3. Indicate he would rather consult and/or do research than teach?
4. Criticize fellow teachers?

*Some of these items would have to be answered after the student received his final grade, particularly number VI-20. The major difficulty here would be administrative, that is, submitting the question or questions to students and having them returned. A suggestion might be to give students the questions during the final examination and ask them to complete and return the form after they have received their final grade. Returns and their representativeness are problematical.

The items presented above can be accepted with some assurance as covering the major effective and ineffective teaching behaviors becuase of their close resemblance to those developed by Konigsburg (1954) and Douglas (1968) in the studies previously described.

The items from these two studies are presented in Appendix 4. Inspection of those lists and comparison with the evaluative items from the present research are assuredly sufficient evidence that these are the teacher behaviors of concern to students with regard to teacher effectiveness. The same behaviors recur in all three lists.

The basic incidents for the three studies have been collected in widely separated parts of the country, from a wide variety of students, and over a long period of time. Despite this disparate collection of critical incidents, there is a quite close resemblance in the reported teacher behaviors.

In general, the three item lists offer a "pool" of behavioral items adequate to assess all of the varied aspects of effective and ineffective teaching behaviors rather than using ratings. This, in turn, offers the opportunity for a research program that will allow determination of exactly which behaviors are related to student achievement and/or attitudes and, further, to what degree each is important in the different student performances.

SUMMARY AND CONCLUSIONS

Research on college teaching effectiveness has been conducted for decades and there is substantial agreement that the results are tenuous, ambiguous, or contradictory. However, there is equal agreement as to the major function of teaching. This is ability to assist students toward agreed upon educational goals.

As Cureton (1951), op. cit., has described it, the ultimate goal of teaching is the best possible life adjustment of persons passing through the educational system insofar as this is a function of that system. Research using this criterion of teacher performance is difficult and expensive to conduct, although not impossible as shown in a study by Hall (1965). In the absence of the practical needs to implement such research, the methodology for the present will need to concentrate on less ultimate criteria and utilize the more immediate, if limited, measures of student achievement and attitudes within the time constraint of single classes or, possibly, a four year college career.

Evaluation of teaching effectiveness immediately entails some measurement of student performance. Such measurement demands some statement of goals, that is, what performance is expected of students. In the foregoing it has been assumed that these goals fall under two broad headings.

The first of these is student achievement. The inference from this goal can largely be described as covering the cognitive domain and refers primarily to student learning. In any specific situation, learning could encompass subject matter particulars, broadened perspectives in an area, relations to other fields of topics, techniques of problem solving, or many other specifications. The basic point is that if teacher performance is to be evaluated, it is essential to provide simultaneous evaluation of the performance of students being taught. In essence, what is the objective or objectives of the teaching and how is student progress toward such objectives to be measured?

The second area of student performance can be summarized under the rather ambiguous term of attitudes. To make this rubric operational there must also be some agreed upon measures. One might measure interest in this course, general opinion of this professor, possibly changes as increased interest in the discipline, or changes in personal perspective. There are numerous possibilities in the area concerning measurements of student attitudes and selection of a specific measurement would depend upon the research problem of interest.

Once the measures of student progress have been agreed upon, the next consideration is a measure of teaching effectiveness. As Barr et. al. (1953), op. cit., have stressed, the need is to determine not who is an effective teacher but what teaching behaviors are effective and ineffective with respect to attaining some established educational and institutional goals. That has been the major effort of this research.

The CIT is specifically designed to determine the effective and ineffective behaviors in any given field of endeavor. It has been applied to college teaching in four separate studies, including the present. The results are in close agreement from these individual efforts in terms of isolating and describing specific, objective behaviors that are noted by students. Moreover, in one study, Douglas (1968), op. cit., the behaviors have been shown to be predictive of student achievement.

This study has taken a broader perspective of the domain of teacher effectiveness and has added some behaviors to those already discovered. It is submitted that the lists of behaviors that have been presented from the four studies form a "pool" of behavioral items that will evaluate teaching effectiveness. They do not pertain to the faculty member performance in terms of research productivity nor effectiveness in staff functions. Those behaviors remain to be investigated but the behaviors do measure what is generally agreed to be the most important teaching duty -- teaching ability. In general, they comprise the important dimensions of teaching and the individual items, a taxonomy of relevant behaviors. They ask the "right questions" mentioned by Gustad (1964), op. cit., Arden (1968), op. cit., and Langlen (1966), op. cit.

If the above formulation, that the behaviors presented do indeed represent a description of the dimensions of effective teaching behavior, the opportunity is available for research to determine the relationship of these behaviors to student performance and the characteristics of both more and less effective teachers and students.

As a first step, it is necessary to regard the reported behaviors as items comprising a psychometric device. The preliminary research would be oriented around determining the characteristics of the device as is customary in the development of psychological tests.

These first studies would gather normative data describing the occurance frequency of the reported behaviors. Next the discrimination values of the individual items, in terms of overall evaluations, would be required and, finally, the intercorrelations of the items and final format reliability determined.

Following such preliminary studies, the vital question of validity would be approached, assuming any indicated revisions of items and format has been accomplished. The needed validation research would be extremely complex and meaningful results could be obtained only from equally complex multivariate studies designed to asses the interactions of the several sets of variables.

Students would comprise one set of variables both as individuals and as a class group. Individual measures would be those available from psychometric devices to test cognitive abilities, personality traits, and interests. Additionally, attitude measures such as those advocated by Pace (1963) and Stern (1963) would tap student perceptions and, inferentially, the school "climate". There is some doubt as to how accurate these latter measures may be, Marks (1967), but since such research is in its infancy, they should be investigated. Class variables such as size, sex ratio, major, and class standing would need to be studied to determine whether or not they influence student perceptions of teacher behaviors.

For teachers, the same general individual measures as those for students would be desirable; however, specific instruments would vary. For example: authoritarianism in teachers has been indicated previously as an important variable in perceptions and evaluations of students. A measure of this variable is desirable in order to assess the basis of teacher behaviors and student perceptions of such behaviors.

The particular course would need to be included in the research design. Here the parameters are somewhat less clear. One specification would be the goals of the course and the concommitant measures of student progress toward such goals. This is an absolute essential. Method of teaching also needs a clear statement, as amount of student participation, work required, and general plan of the course. Here parameters would probably be discovered as teaching behaviors are used as an independent variable. In particular, the class as a "social system" is an area that needs investigation as noted by Gibb (1955), op. cit. There is relatively little research information concerning the social aspects of teacher-student interactions.

Using all these sets of variables dictate the requirement of large and complex studies but these are vital since there are little or no research data describing the interactions of the several possible and important variables. Most of the reported research has used a limited number of variables and different ones, in consequence, it is necessary to proceed on the assumption that the effects of all variables are unknown. The need for such studies has been expressed previously, in particular, Metzel (1960), op. cit., and Anderson and Hunka (1963) have presented quite explicit conceptualizations of the need.

The statements above have sketched in broadest outlines the needed research. The interested reader is referred to McKeachie (1968) and Barr, et. al., (1953), op. cit., for some detailed formulations of research requirements. However, studies designed as above should eventually indicate a set of somewhat limited teacher behaviors that are crucial to both student achievement and attitudes. The practical situation of evaluating teacher performance requires the use of an instrument that can be used by students without demanding too much in the way of their time or class time. With this consideration in mind, future research should aim at determining the relative importance of the various behaviors with regard to student learning and attitudes and these would comprise the evaluation instrument, the minimum needed to assess teacher performance adequately.

Concurrent with the research above, data would be developed describing the personal correlates of effective and ineffective classroom teaching behaviors for the individual teachers used as subjects in the several research studies. Such data would be invaluable in future selection and training of college faculty members. At present such selection and training are on a hit-or-miss basis, since no adequate performance criteria have been available. Assuming the use of the effective and ineffective behaviors derived from the critical incident studies as criteria, both selection and training of college teachers can be developed on a scientific basis with beneficial results for all concerned.

In conclusion, it is submitted that the behaviors described in the four critical incident studies have resulted in a taxonomy of teaching behaviors that can serve as criteria for college teaching evaluation for both research and administrative purposes. Further research is needed to refine and clarify the several behaviors to develop practical evaluation instruments and to determine the correlates of student achievement and attitudes along with those of the effective college teacher.

BIBLIOGRAPHY

Anikeef, A.M. Factors Affecting Student Evaluation of College Faculty Members. Journal of Applied Psychology, 1953, 37 (6), 458-460.

Arden, E. Faculty as Teachers. Educational Forum, 1968, 32, 447-452.

Barr, A.S. The Measurement and Prediction of Teaching Efficiency: A summary of investigation. Journal of Experimental Education, 1948, 16, 203-283.

Barr, A.S., Bechdolt, B.V., Coxe, W.W., Gage, N.L., Orleans, J.S., Remmers, H.H., and Ryans, D.C. Report of the committee on the Criteria of Teacher Effectiveness. Review of Educational Research, 1953, 22, 238-263.

Barr, A.S. and Jones, R. The Measurement and Prediction of Teacher Efficiency. Review of Educational Research, 1958, 28, 256-264.

Bendig, A.W. The Relation of Level of Course Achievement to Students' Instructor and Course Ratings in Introductory Psychology. Educational and Psychological Measurement, 1953a, 13, 437-448.

Bendig, A.W. Student Achievement in Introductory Psychology and Student Ratings of the Competence and Empathy of Their Instructors. Journal of Psychology, 1953b, 3b, 427-433.

Bendig, A.W. A Statistical Report on a Revising of the Miami Instructor Rating Sheet. Journal of Educational Psychology, 1952, 43, 423-429.

Bloom, B.S. (Ed.) Taxonomy of Educational Objectives. Handbook I: Cognitive domain. New York: Longmans, 1956.

Borg, W.R., and Hamilton, E.R. Comparison Between a Performance Test and Criteria of Instructor Effectiveness. Psychological Reports, 1956, 2, 111-116.

Bousfield, W.A. Students' Ratings of Qualities Considered Desirable in College Professors. School and Society, 1940, 51, 253-256.

Coffman, W.E. Determining Students' Concepts of Effective Teaching From Their Ratings of Instructors. Journal of Educational Psychology, 1954, 45, 277-286.

Cook, W.W. and Leeds, C.H. Measuring the Teaching Personality. Education and Psychological Measurement, 1947, 7, 399-410.

Cosgrove, D.J. Diagnostic Rating of Teacher Performance. Journal of Educational Psychology, 1959, 50 (5), 200-204.

Crannell, C.W. A Preliminary Attempt to Identify the Factors in Student - Faculty Evaluation. Journal of Psychology, 1953, 36, 417.

Crawford, P.L. and Bradshaw, H.L. Perception of Characteristics of Effective University Teachers: A Scaling Analysis. Education and Psychological Measurement, 1968, 28 (2), 1079-1085.

Domas, S.J. Report of an Exploratory Study of Teacher Competence. Cambridge, Mass.: Peabody House, 1950. CIT Study.

Domas, S.J., and Tiedeman, D.V. Teacher Competence: An annotated bibliography. Journal of Experimental Education, 1950, 19, 101-218.

Douglass, L.C. Measures of Teacher Evaluation As Related to Student Achievement. M.S. Thesis, University of Tennessee, 1968.

Downie, N.M. Student Evaluation of Faculty. Journal of Higher Education, 1952, 23, 495f.

Eble, K.E. Special Report: Project to Improve College Teaching. Washington, D.C.: American Association of University Professors (undated).

Eells, W.C. College Teachers and College Teaching: An Annotated Bibliography. Atlanta, Georgia: Southern Regional Education Board, 1957.

Flanagan, J.C. The Critical Incident Technique. Psychological Bulletin, 1954, 51, 327-358.

Gibb, C.A. Classroom Behavior of the College Teacher. Educational and Psychological Measurement, 1955, 15, 254-263.

Gustad, J.W. Policies and Practices in Faculty Evaluation. Educational Record, 1961, 42, 194.

Hall, V.C. Former Student Evaluation As a Criterion for Teaching Success. Journal of Experimental Education, 1965, 34, 1-19.

Heilman, J.D. and Armentrout, W.D. The Rating of College Teachers on Ten Traits By Their Students. Journal of Educational Psychology, 1936, 27, 197.

Holland, J.B. The Image of the Instructor as it is Related to Class Size. Journal of Experimental Education, 1954, 23, 171-177.

Isaacson, R.L., McKeachie, W.J., Milholland, J.E. Correlation of Teacher Personality Variables and Student Ratings. Journal of Educational Psychology, 1963, 4 (2), 110-117.

Jensen, A.C. Determining Critical Requirements for Teachers. Journal of Experimental Education, 1951, 20, 79-85.

Konigsburg, D. Development and Preliminary Evaluation of an Instructor Checklist Based on the Critical Technique. Doctoral Dissertation Series, 1954, #7937.

Krasner, L. Behavior Therapy in Mussen, P.H. and Rosensweig, M.R. Annual Review of Psychology. Palo Alto, California: Annual Reviews, Inc., 1971.

Lacognata, A.A. University Extension Faculty and Student Role Expectation: An Empirical Analysis. Journal of Experimental Education, 1964, 33, 107-120.

Lifson, K.A. Errors in Time - Study Judgements of Industrial Work Pace. Psychological Monographs, 1953, 67, 355.

Lovell, G.D. and Haner, C.F. Forced-Choice Applied to College Faculty Rating. Educational and Psychological Measurement, 1955, 15, 291-304.

Marks, E. Personality and Motivational Factors in Responses to an Environmental Description Scale. Atlanta, Georgia: Georgia Institute of Technology, 1967.

Maslow, A.H. and Zimmerman, W. College Teaching Ability, Scholarly Activity, and Personality. Journal of Educational Psychology, 1956, 47, 185-89.

McGrath, E.J. Characteristics of Outstanding College Teachers. Journal of Higher Education, 1962, 33, 148.

Mischel, W. Personality and Assessment. New York: John Wiley and Sons, 1968.

Moore, J.E. A Further Study of the Annoying Habits of College Professors. Journal of Abnormal and Social Psychology, 1937, 32, 368-375.

Morsh, J.E. and Wilder, E.W. Identifying the Effective Instructor: A Review of the Quantitative Studies, 1900-1952. San Antonio, Texas: Air Force Personnel and Training Research Center, 1954.

Pace, C.R. College and University Environment Scales: Preliminary Manual. Princeton, N.J.: Educational Testing Service, 1963.

Prien, E.P. Assessments of High-level Personnel: V. An Analysis of Interviewers' Prediction of Job Performance. Personnel Psychology, 1962, 15, 319-334.

Prien, E.P. and Powell, D.R. A Study of the Director's Functions. Journal of the American Society of Training Directors, 1961, 15, 12-17.

Rayder, N.F. College Student Ratings of Instructors. Journal of Experimental Education, 1968, 37 (2), 76-81.

Remmers, H.H. Rating Methods in Research on Teachers. Gage, N.L. (Ed.) Handbook of Research on Teaching, Chicago: Rand McNally, 1963.

Remmers, H.H. Reliability and Halo Effect of High School and College Students' Judgements of Their Teachers. Journal of Applied Psychology, 1934, 18, 619-630.

Ronan, W.W. The Relationship of Personality and Aptitude Factors. Experimental Publication System, 1969, 3, 76.

Ronan, W.W. A Factor Analysis of Eleven Job Performance Measures. Personnel Psychology, 1963, 16, 255-267.

Ronan, W.W. and Prien, E.P. Perspectives on the Measurement of Human Performance. New York: Appleton - Century - Crofts, 1971.

Ronan, W.W. and Prien, E.P. Toward a Criterion Theory: A Review and Analysis of Research and Opinion. Greensboro, N.C.: The Richardson Foundation, 1966.

Rosenthal, R. Experimenter Effects in Behavioral Research. New York: Appleton - Century Crofts, 1966.

Rush, C.A. A Factorial Study of Sales Criteria. Personnel Psychology, 1953, 6, 9-24.

Russell, H.E. and Bendig, A.W. An Investigation of the Relation of Student Ratings of Psychology Instructors to Their Course Achievement When Academic Aptitude Is Controlled. Educational and Psychological Measurement, 1953, 13, 626-635.

Ryans, D.G. Characteristics of Teachers. Washington, D.C.: American Council on Education, 1960.

Ryans, D.G. Notes on the Rating of Teacher Performance. Journal of Educational Research, 1954, 47, 695-703.

Smalzried, N.T. and Remmers, H.H. A Factor Analysis of the Purdue Rating Scale for Instructors. Journal of Educational Psychology, 1943, 34, 363-367.

Smit, J.A. A Study of the Critical Requirements for Instructors of General Psychology Courses. unpublished doctoral dissertation, University of Pittsburgh, 1951.

Solomon, D. Teacher Behavior Dimensions, Course Characteristics, and Student Evaluations of Teachers. American Educational Research Journal, 1966, 3, 35.

Solomon, D., Rosenberg, L. and Bezdek, W.E. Teacher Behavior and Student Learning. Journal of Educational Psychology, 1964, 55 (1), 23-30.

Stern, G.G. Characteristics of the Intellectual Climate in College Environments. Harvard Educational Review, 1963, 33, 5-41.

Taft, M.J. A Systematic Method for Evaluating Teachers. Journal of Engineering Education, 1969, 59, 852-857.

Taylor, C.W., Price, P.B., Richards, J.M., and Jacobsen, J.L. An Investigation of the Criterion Problem for a Medical School Faculty. Journal of Applied Psychology, 1964, 48, 294-301.

Turner, W.W. Dimensions of Foreman Performance: A Factor Analysis of Criterion Measures. Journal of Applied Psychology, 1960, 44, 216-223.

Weaver, C.H. Instructor Rating by College Students. Journal of Educational Psychology, 1960, 51, 21-25.

Webb, S.C. Student Perceptions of Instructor Teaching Goals In Correspondence with Instructor Self-Ratings. Georgia Tech: Office of Evaluation Studies, Research Memorandum 67-5, May 1967.

Wherry, R.J. Control of Bias in Rating (Sub-project 2) Instructor Rating Scales. Washington, D.C.: Personnel Research Section, AGO, U.S. Department of the Army, 1950.

APPENDIX 1

TABULATION OF EFFECTIVE AND INEFFECTIVE INCIDENTS BY CATEGORY AND SUB-CATEGORY

EFFECTIVE INCIDENTS*

I. Personal Relationships with Students (379)

1. Knows each student by name (48)
2. Mingles with students before and after class (8)
3. Holds social events for his students (7)
4. Gives (encourages students to ask for) advised and assistance regarding personal problems or goals (41)
5. Discusses (answers questions about) extra-class issues with students (6)
6. Compliments a student on a good response (3)
7. Explains answers in detail to all (encourages) questions asked by students (51)
8. Treats all students fairly regardless of sex, race, etc. (4)
9. Holds special problems sessions or allots class time for questions (39)
10. Gives individual help, in class or office, without hesitation (encourages students to ask for) (157)
11. Miscellaneous (15)

II. Classroom Administration (80)

1. Extends time limit (changes dates) on assignments and quizzes (6)
2. Informs class of days he will be absent or changes in plans (4)
3. Lecture begins and ends on time (4)
4. Distributes or details a study plan the first week of class outlining the course requirements (11)
5. Follows course syllabus or lecture (outline) on schedule (11)
6. Gives examples of quiz items or what to expect on quiz in class (6)
7. Keeps old quiz questions on file for student inspection (10)
8. Requires and grades homework (21)
9. Grades papers or quizzes promptly (7)

III. Student Participation (124)

1. Permits students to determine part or all of course content, class policy (20)
2. Improves his course by making changes based on criticism requested from students (9)
3. Schedules quiz at convenience of class majority (11)

*Number of incidents in parentheses.

4. Encourages group discussion and differences of opinion (65)
5. Seeks feedback from students, in particular, understanding of material (19)

IV. Classroom Presence (65)

1. Makes dramatic gestures and comments to emphasize important points (3)
2. Speaks in a clear, distinct manner, diction and grammar or both (3)
3. Uses humor that stimulates class interest and attendance (38)
4. Lectures without relying solely (reading) on notes or text (3)
5. Uses language that students can understand (not talk down) (14)
6. Personal appearance (4)

V. Organization and Presentation of Material (412)

1. Begins each class with a review of previous work (6)
2. Stresses important points, general concepts in teaching (21)
3. Puts important information on the board in clear, concise manner (16)
4. Uses current and pertinent examples and illustrations to explain material (40)
5. Shows relevance of subject to the "real world", the students' major, and/or students' outside interests or future (107)
6. Asks questions in class (stimulates thought, interest) (20)
7. Admits answer he does not know/provides answer next lecture (7)
8. Lectures reinforce textbook (5)
9. Supplements course (book) by using outside references and materials (30)
10. Distributes hand-outs and/or copy of class notes to supplement course (quiz) (32)
11. Supplements lectures with visual aids (blackboard) (34)
12. Provides field trips (4)
13. Invites guest lecturers who are specialists on course topic (6)
14. Explains (works out) answers to quiz, homework and class problems (35)
15. Does not rigidly follow book or notes in his lectures (well-organized and prepared) (11)
16. Uses department (personal) experiments, projects, or work to stimulate student interest (19)
17. Has full (or personal) command of subject matter (10)
18. Reviews material before a quiz or assignment (study guides, assigns similar problems) (11)
19. Pace of lectures can be followed (to take notes) (8)

VI. Evaluation of Student Performance (222)

1. Tests are based upon lectures, text, and/or homework (announced and/or relevant) (70)
2. Tests require knowledge of principles rather than memory alone (6)
3. Gives take-home final and/or open book quiz (use of class notes) (8)
4. Schedules quizzes at regular intervals (8)
5. Writes comments (reviews) on returned papers and quizzes (7)
6. Students with high average are execused from the final exam (11)
7. Students with low average are permitted to do extra work (test) (13)
8. Disregards the lowest test score of each student (optional test) (11)
9. Permits make-up tests at individual convenience (13)
10. Takes into account class participation, application, and/or effort in assigning final grades (11)
11. Curves grades on the basis of class distribution (21)
12. Does not penalize for class absence or tardiness (accepts excuses) (17)
13. Reviews test scores and/or changes grade if warranted (16)
14. Grades papers himself rather than employing a student grader (4)
15. Adequate time to complete tests (4)
16. Miscellaneous (1)

INEFFECTIVE BEHAVIORS

I. Personal Relationships with Students (264)

1. Shows favoritism toward some students (athletes, apple-polishers, reciters, etc.) (14)
2. Singles out some students as inferior (discriminates) (17)
3. Ridicules or embarrasses students (question or performance) (59)
4. Loses control of emotions in dealing with students (shouts, curses, etc.) (20)
5. Harasses students during tests, lab work, reports, and answering questions (5)
6. Demoralizes students by threatening punitive actions (11)
7. Does not accept legitimate excuses or explanations (14)
8. Does not know (or attempt to) students on a personal basis (e.g. by name) (5)
9. Hesitates or refuses to answer questions (inadequate answers) (61)
10. Hesitates or refuses to help students (class or office) (50)
11. Dogmatic and inflexible (belittles students in general) (9)

II. Class Administration (149)

1. Meets class irregularly or not at all (special sessions). Leaves lab (43)
2. Frequently comes to class late (35)
3. Permits classroom disturbances, lack of attention (4)
4. Consistently lectures overtime (13)
5. Fails to state objectives and overall purpose of course (3)
6. Makes false statements regarding course requirements and what is expected of students (25)
7. Monopolizes student time with excessive or irrelevant assignments (21)
8. Gives no exam before drop date (5)

III. Student Participation (37)

1. Does not permit class discussion of scheduling quizzes or assignment due dates (2)
2. Does not permit or encourage class discussion of material or opinion (35)

IV. Classroom Presence (338)

1. Objectionable dress, manners, and appearance (11)
2. Displays nervousness, ill-at-ease, when talking (e.g., paces floor, easily flustered) (8)
3. Talks or presents material too rapidly (22)
4. Lectures in a rambling, disorganized fashion (8)
5. Speaks inaudibly and/or mumbles (55)
6. Lectures in a monotone (51)
7. Difficulty in speaking English (27)
8. Does not look at students during lecture (16)
9. Reads majority of lecture from book or notes instead of just referring to them (103)
10. Uses profane language constantly (7)

V. Organization and Presentation of Material (222)

1. Does not cover all of the course requirements (8)
2. Wastes class time on trivial detail and/or subjects unrelated to course objectives (51)
3. Repeats material to the point of monotony (same lecture - different course) (7)
4. Lectures above students' level of understanding (15)
5. Unable to work problems or answer questions (60)
6. Gives erroneous information (7)
7. Does not or refuses to explain course material (9)
8. Forces students to shoulder burden of gaining subject matter knowledge (28)
9. Lectures do not contain any material that is not fully explained in the book (1)
10. Hurries through course schedule without regard for student understanding of material (7)
11. Lectures consist of copious blackboard notes (2)
12. Stresses theory without explaining applicability (15)
13. Unprepared for class (12)

VI. Evaluation of Student Performance (342)

1. Tests students on material that was not assigned (40)
2. Tests do not include material emphasized in class (covered) and/or in reading assignments (47)
3. Tests either exceed the difficulty level of the material or are too easy (9)

4. Tests require memorization rather than demonstrating knowledge of principles or own ideas (22)
5. Time to complete tests is inadequate (6)
6. Uses same test questions every year (some students obtain tests) (31)
7. Refuses to (does not) discuss or explain returned assignments, tests, projects or grades (grading system) (30)
8. Does not comment on returned papers (tests) (12)
9. Returns papers (tests) late or not at all (22)
10. Does not grade quizzes or assignments (collect) (9)
11. Grades on classroom participation only (1)
12. Grades on irrelevant characteristics (dress, major, sex, own biases, etc.) (15)
13. Grades on class attendance (34)
14. Grades on final exam only (10)
15. Grade is not in accord with test scores (10)
16. Does not give credit for partially correct answers (13)
17. Student on borderline is given lower grade (21)
18. Passes and/or fails or gives grades to a predetermined percentage of students in class (or large percentage) (23)
19. Does not curve grades (14)
20. Does not check accuracy of student grader (3)
21. Make-up tests are made excessively difficult (21)
22. Gives no quizzes and/or final (3)
23. Refuses to change incorrect grade (8)

VII. Interest in Job of Teaching (19)

1. Makes derogatory comments about teaching (3)
2. Belittles value of course he is teaching (6)
3. Criticizes fellow teachers (2)
4. Primary interest is consulting or research (8)

APPENDIX 2

NUMBER OF INCIDENTS REPORTED BY MAJOR

NUMBER OF INCIDENTS REPORTED BY MAJOR

	Major	Number Effective Incidents	Number Ineffective Incidents
1.	Aerospace Engineering	52	49
2.	Architecture	38	31
3.	Biology	25	22
4.	Business Administration	1	7
5.	Ceramic Engineering	3	13
6.	Chemical Engineering	64	33
7.	Chemistry	24	24
8.	Civil Engineering	59	56
9.	Electrical Engineering	61	72
10.	English	3	1
11.	History	4	3
12.	Industrial Engineering	108	103
13.	Industrial Management	550	492
14.	Information Sciences	8	14
15.	Mathematics	51	41
16.	Mechanical Engineering	55	72
17.	Mechanics	9	8
18.	Nuclear Engineering	0	1
19.	Physics	30	30
20.	Psychology	5	7
21.	Sociology	4	2
22.	Textile Engineering	34	27

Rho = 0.947, Effective vs. Ineffective

CATEGORY - SUB-CATEGORIES

Major	Effective I Personal Relationships With Students 1	2	3	4	5	6	7	8	9	10	11	Total	Ineffective I Personal Relationships With Students 1	2	3	4	5	6	7	8	9	10	11	Total
Aerospace Engineering	3	1		2			3		5	6		20	1		1						6	2		10
Architecture	1		1	2			1			8	1	14	1	3	2	2	1		1			1		11
Biology							2	2		2		6	2		3			1			3			9
Business Administration												0									2			2
Ceramic Engineering												0												0
Chemical Engineering	3			3			2			9		17			3						3	2	1	9
Chemistry				1			1			3		5		1							2	2		5
Civil Engineering	2		2	2			3		2	7	1	19			1	1			1		4	2	1	10
Electrical Engineering	1						1		2	4		8		1	1	1			1		2	1	1	8
English												0												0
History	1											1				1								1
Industrial Engineering	5			7	1		4		2	11		30	1	2	7	1		3	1		4	3		22
Industrial Management	19	3	4	15	5	3	14	1	16	67	8	155	7	5	27	8	4	3	9	3	17	27	4	114
Information Sciences							1			2	1	4									1			1
Mathematics	3	2		2			2		3	4		16		2	1			1		1	2	3		10
Mechanical Engineering				1			8		3	5	1	18			5	2					3	1	1	12
Mechanics												0	1											1
Nuclear Engineering												0												0
Physics	1			4			5		1	3		14		1		1			1	1	3	3	1	11
Psychology									1	3	1	5	1	1		1								3
Sociology	1											1				1								1
Textile Engineering	2						2	1	3	6		14		1		1		1						3

APPENDIX 3

CATEGORIES OF INCIDENTS REPORTED BY MAJOR

CATEGORY - SUB-CATEGORIES

Major	Effective II Class Administration										Ineffective II Class Administration								
	1	2	3	4	5	6	7	8	9	Total	1	2	3	4	5	6	7	8	Total
Aerospace Engineering			1					1		2	2	4		2		1	1		10
Architecture	1							1		2		2				4		1	7
Biology					1	1				2	1						1		2
Business Administration										0							1		1
Ceramic Engineering										0		1							1
Chemical Engineering				1		1	1	1		4	1	1				2			4
Chemistry	1								1	2		1		1		1			3
Civil Engineering								1		1	1	3		2			1		7
Electrical Engineering				2	2		4	2		10	2	3				1			6
English										0									0
History				1	1					2							1		1
Industrial Engineering	1	1		1	1			1	2	7	3	2			1	1	2		9
Industrial Management	2	1	1	3	4	4	5	8	3	31	24	11	2	6	1	9	10	3	66
Information Sciences										0					1				1
Mathematics	1		1	1	1	1		2	1	8	1	1	1	2		3	2	1	11
Mechanical Engineering		1	1	1		1		1		5	2						1		3
Mechanics	1				1					2		1							1
Nuclear Engineering										0									0
Physics								2		2			1			1			2
Psychology										0	1						1		2
Sociology										0									0
Textile Engineering										0	2								2

CATEGORY - SUB-CATEGORIES

	Effective III Student Participation						Ineffective III Student Participation		
Major	1	2	3	4	5	Total	1	2	Total
Aerospace Engineering		1		1		2	1		1
Architecture				3	3	6		2	2
Biology	1			3		4		1	1
Business Administration						0			0
Ceramic Engineering						0			0
Chemical Engineering				2	2	4		1	1
Chemistry		1		1		2		1	1
Civil Engineering	1		1	3	2	7			0
Electrical Engineering	2			2		4		2	2
English				2		2			0
History						0			0
Industrial Engineering	2	3	1	3	5	14		3	3
Industrial Management	13	2	7	29	6	57	1	19	20
Information Science		1				1		2	2
Mathematics			1	4	1	6		1	1
Mechanical Engineering				3		3			0
Mechanics						0			0
Nuclear Engineering						0			0
Physics				3		3			0
Psychology						0		1	1
Sociology						0			0
Textile Engineering				1		1			0

CATEGORY - SUB-CATEGORIES

Major	Effective IV Classroom Presence							Ineffective IV Classroom Presence										
	1	2	3	4	5	6	Total	1	2	3	4	5	6	7	8	9	10	Tota
Aerospace Engineering			2		1		3	1				2	3	2	3	5		16
Architecture			3				3	1		1		2				1		5
Biology			1				1						2	1	2	3		8
Business Administration							0									1		1
Ceramic Engineering							0											0
Chemical Engineering	1		4	1			6		1	1	1	2	1	2		2		10
Chemistry			2		1		3	1						1		2		4
Civil Engineering			4				4	2				3	1	3		5	1	15
Electrical Engineering			2		4		6	1		1	1	2	8	1	5	13		32
English							0											0
History							0											0
Industrial Engineering			3			1	4		2	1	1	5	3	2	3	7	1	25
Industrial Management	1	2	10		6		19	2	3	13	4	32	19	7	21	34	3	138
Information Science							0	2							2			4
Mathematics			1	1			2			1		1	3		1	5	1	12
Mechanical Engineering			3	1			4			3	1	1	4	6	3	7		25
Mechanics						1	1											0
Nuclear Engineering							0											0
Physics							0		1				3	2	1	4		11
Psychology							0											0
Sociology							0											0
Textile Engineering	1					1	2						1		1	3		5

CATEGORY - SUB-CATEGORIES

Effective V
Organization & Presentation of Material

Major	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Total
Aerospace Eng.			2	1	6	1			1		2			2	2		1	1		19
Architecture		1			2				3					1						7
Biology			2	1						2	1					1		1		8
Business Admin.					1															1
Ceramic Engr.				1																1
Chemical Engr.		5	1		6	1	1		2	2	1			2	1			1		23
Chemistry	1			1	2									1				1		6
Civil Engr.		2		2	4					1	2			2	2	1	2	1	1	20
Electrical Engr.		2		2	7	2			3	2	1			3						23
English															1					1
History																				0
Industrial Engr.	1	3	2	5	11	2			2	3	1	1		2				1	3	37
Industrial Mgmt.	12	6	5	19	50	12	2	2	16	19	6	2	3	12	5	13	5	1	2	184
Information Sci.			1	1																2
Mathematics			1	1	3		1	1	1		2			3			1			14
Mechanical Engr.	1		2	3	3		1				4	1		1		1	2			19
Mechanics					1			1		1				2			1			6
Nuclear Engr.																				0
Physics		1	1		1					1	1			1		1	1			8
Psychology																				0
Sociology					1				1		1									3
Textile Engr.					5	1								1		1			1	9

CATEGORY - SUB-CATEGORIES

Major	Effective V Organization of Material 1	2	3	4	5	6	7	8	9	10	11	12	13	Total
Aerospace Engr.	1				4		1	2				1		9
Architecture		1	1		1						1			4
Biology		3			2									5
Business Admin.		2												2
Ceramic Engr.		2												2
Chemical Engr.		1	3	1				1	3					9
Chemistry			1			1								2
Civil Engr.	1	4	1	1	3	1	1	1		1		3		17
Electrical Engr.				1	1	3			2				1	8
English														0
History				1										1
Industrial Engr.		1		1	6	1	1	2		2	1	1	1	17
Industrial Mgmt.	4	18	3	6	24	1	1	7		4	1	5	8	82
Information Sci.				2	1	1	1					1		6
Mathematics		2			3			1						6
Mechanical Engr.				2	3	1	1	3				2	1	13
Mechanics		3											1	4
Nuclear Engr.														0
Physics		1			3			2						6
Psychology		1												1
Sociology														0
Textile Engr.		3		1			1		1					6

CATEGORY - SUB-CATEGORIES

Effective VI

Major	Evaluation of Student Performance																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
Aerospace Engr.	1					2	1				1	1					6
Architecture			1	1	1		2		1								6
Biology	2	1			1												4
Business Admin.																	0
Ceramic Engr.								1			1						2
Chemical Engr.	2	1		1			2	1						1			8
Chemistry	1		1							1			1	2			6
Civil Engr.	2				1		1	1			1	2		1			9
Electrical Engr.	4		1				1				2			2			10
English																	0
History	1																1
Industrial Engr.	7	1		1	1	1		2	2	2	1	4	1				23
Industrial Mgmt.	35	1	5	3		8	5	5	8	6	11	7	2	2	1	5	104
Information Sci.					1												1
Mathematics	1			1					1		1	1					5
Mechanical Engr.	2				1						1	1	1				6
Mechanics																	0
Nuclear Engr.																	0
Physics	1	2															3
Psychology																	0
Sociology																	0
Textile Engr.	5							2	1								8

CATEGORY - SUB-CATEGORIES

Ineffective VI
Evaluation of Student Behavior

Major	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total
Aerospace Engr.	3							2		1		1	1		1			2					1	1	13
Architecture	1	1					1		1					1				1						1	7
Biology	3	2							2			1													8
Business Admin.									1																1
Ceramic Engr.		1																							1
Chemical Engr.	4	4		1					1	1		1		1					1						14
Chemistry	1	1					2	1					1		1			1	1	1			1		11
Civil Engr.	1				1		3	1	2	1		1	1					1	1				1		14
Electrical Engr.	2	2		1	3		3	1	2	1			2			2			1				1		21
English								1																	1
History																1									1
Industrial Engr.	3	4	2	2			2	3	2				4	2	1	1	1		2		1	1	1		32
Industrial Mgmt.	20	19	3	14	2	1	9	3	8	4	1	6	14	2	6	5		6	4	2	1	1	1	2	134
Information Sci.							1																		1
Mathematics	1						2		1	1			3	1				1					2		12
Mechanical Engr.	2	4					2					1	5	1				3	1						21
Mechanics							1					1		1											3
Nuclear Engr.												1													1
Physics				1														1						1	3
Psychology												1						1							2
Sociology																									0
Textile Engr.		4	1									1	3		1			1				1			12

CATEGORY - SUBCATEGORIES

Major	Ineffective VII Interest in the Job of Teaching 1	2	3	4	Total
Aerospace Engineering					0
Architecture		1		1	2
Biology				1	1
Business Administration		1			1
Ceramic Engineering					0
Chemical Engineering					0
Chemistry		1			1
Civil Engineering					0
Electrical Engineering				1	1
English					0
History					0
Industrial Engineering	2		2		4
Industrial Management		1		3	4
Information Sciences					0
Mathematics					0
Mechanical Engineering	1				1
Mechanics					0
Nuclear Engineering					0
Physics					0
Psychology					0
Sociology					0
Textile Engineering		1			1

APPENDIX 4

BEHAVIORAL EVALUATION STATEMENTS DERIVED FROM OTHER CRITICAL INCIDENT STUDIES

The questions below are from the study by Konigsburg (1954) discussed previously:

DURING THE PAST THREE CLASS PERIODS, did the instructor:

		YES	NO
1.	Use term which you did not understand?	____	____
2.	Speak so rapidly that you could not take notes?	____	____
3.	Use incidents which occurred in or out of class to clarify points?	____	____
4.	Speak in such a low tone that you were unable to understand him?	____	____
5.	Read continually from lecture notes instead of only referring to them?	____	____
6.	Start to explain a new concept before completing his explanation of the previous one?	____	____
7.	Present too much material for you to learn in a class period?	____	____
8.	Refuse to answer students' questions?	____	____
9.	Speak in a monotone, seldom raising or lowering his voice?	____	____
10.	Leave out steps or present steps out of sequence in explaining a concept?	____	____
11.	Tell about his own work or an experiment he has participated in which was related to class topic?	____	____
12.	Review or discuss during class period all or part of an article related to class topic?	____	____
13.	Spend part of each class period summarizing material from the previous class period?	____	____
14.	Lecture on material which, in order to be understood, required training you have not yet had?	____	____

DURING THE PAST THREE CLASS PERIODS, did the instructor:

		YES	NO
15.	Describe incidents in the life of an individual (ex., present a case history) which were related to class topic?	____	____
16.	Have every student prepare questions about difficult materials for review at a later time?	____	____
17.	Fail to give examples to clarify basic points in lecture?	____	____
18.	Give lecture on materials which did not help you to understand the material in the textbook?	____	____
19.	Describe an experiment which was related to class topic?	____	____
20.	Show models or apparatus which helped you to understand the subject matter?	____	____
21.	Show movies which helped you to understand the subject materials being discussed?	____	____
22.	Put drawings, diagrams, and charts on blackboard?	____	____
23.	Put outline on blackboard which indicated each step to be taken in order to solve a problem?	____	____
24.	Select students from class as subjects or use all students in class as subjects?	____	____
25.	Illustrate topic being discussed by using diagrams, charts, and graphs which were not on blackboard?	____	____
26.	Give incorrect answers to questions?	____	____
27.	Have students given demonstrations and talks?	____	____
28.	Present problem and ask members of class to give their reactions to it?	____	____
29.	Ask the class questions?	____	____

DURING THE PAST THREE CLASS PERIODS, did the instructor:

		YES	NO
30.	Fail to provide time for answering questions about parts or all of previous assignment?	____	____
31.	Give an ungraded check test?	____	____
32.	Permit students to talk among themselves at the same time that he was talking to class or a student was reciting?	____	____
33.	Permit student to monopolize class time by arguing a particular point?	____	____
34.	Tell a student he would not be able to master material?	____	____
35.	Fail to adhere to schedule?	____	____
36.	Come to class late?	____	____
37.	Announce times, other than during class period, when he would be available to discuss class problems?	____	____
38.	Listen to all student comments and try to show the relationship between the comments and the subject matter?	____	____
39.	First refer to the textbook or to his notes before answering students' questions?	____	____
40.	Pace the floor, rustel papers, or in other ways, draw your attention to him instead of to the meaning of what he was saying?	____	____
41.	Turn his back to the class while lecturing?	____	____
42.	Make a remark or comment about a student which was followed by laughter from the other students?	____	____
43.	Look out the window or at floor while lecturing?	____	____

The statements below are from the Douglas (1968) study discussed previously.

DIRECTIONS: Recall the teaching performance this quarter of the professor you are considering. Check each of the performance statements below which you definitely observed. If you do not immediately recall the performance, then do not check the statement.

E 1. Lectured fluently without aid of notes.

E 2. Used leading questions to force a student to answer his own questions.

E 3. Pointed out relationships between his field and other fields of study.

E 4. Kept details to a minimum, concentrating on broad aspects.

I 5. Met class irregularly or not at all.

E 6. Made frequent written homework assignments which effectively prevented students from falling behind in their work.

I 7. Criticized students' questions or reactions repeatedly.

I 8. Failed to provide adequate explanation or clarification in his lectures.

I 9. Made little or no attempt to provide knowledge of results on assignments or tests.

I 10. Belittled the course he was teaching.

I 11. Made reading assignments in the text but never included the text in his lectures.

I 12. Tested students on material that had not been assigned.

E 13. Created an effective atmosphere for learning by the informal, relaxed manner in which the class was conducted.

I 14. Lectured in a monotone.

E 15. Brought in outside resource people that enhanced the understanding of the subject.

E 16. Gave occasional "pop" quizzes which resulted in helping the student stay abreast of class work.

I 17. Repeatedly called on certain students for class discussion.

E 18. Introduced humor to stimulate class interest.

E 19. Forced students to qualify, explain, or justify statements and assertions they made in class.

E 20. Demonstrated the importance and significance of his subject matter.

I 21. Displayed nervousness or fear when lecturing.

I 22. Monopolized student study time with outside assignments.

I 23. Hesitated or refused to answer students' questions during class.

E 24. Clearly stated the purposes and objectives of the course.

E 25. Adjusted his pace to the needs of the class.

E 26. Lectured in a manner which facilitated note-taking.

I 27. Required students to memorize lists of facts without demonstrating their relevance.

E 28. Presented examples or personal experiences to illustrate lectures.

E 29. Supplemented the text by lecturing from other sources and relating these sources to the text.

E 30. Demonstrated the practical application of the course content.

I 31. Introduced unfamiliar words or technical terms without clarification.

I 32. Repeated material to the point of monotony.

E 33. Increased student alertness by use of skillful and repeated oral questioning in class.

I 34. Required excessive and unnecessary memorization.

E 35. Prepared the student for difficulties that might be encountered on an assignment.

E 36. Used class discussions to bring out contrasting views.

I 37. Administered tests infrequently, forcing students to cover too much material for a single test.

E 38. Allowed students to choose their own topics for assigned papers, reports, or projects.

E 39. Extended his office hours in order to further assist students.

I 40. Consistently lectured on material never covered on exams.

E 41. Demonstrated tolerance toward students' ideas even when they conflicted with lectures or course materials.

E 42. Arranged "help sessions" for students at their request.

E 43. Devised projects requiring students to prepare and present speeches or problems relevant to course content.

I 44. Came to class unprepared.

I 45. Lectured on subjects unrelated to the course.

E 46. Required students to arrive at their own conclusions on class discussions or problems.

E 47. Summarized material and showed relationships in a manner which aided retention.

E 48. Prepared hand-out sheets to complement lectures.

E 49. Differentiated between significant and non-significant material.

E 50. Answered students' questions adequately.

E 51. Referred to current happenings in his field which were not mentioned in the text.

E 52. Learned students' names or showed other indication of recognition.

E 53. Requested and obtained students' questions and reactions.

E 54. Set aside some class time to review homework assignments.

I 55. Spoke indistinctly.

I 56. Refused to consider alternative approaches to problems or issues.

I 57. Lectured in a manner which failed to hold class attention.

I 58. Permitted classroom distractions to go unchecked.

I 59. Demoralized the class by announcing his intention of giving all low grades.

E 60. Assigned text or outside reading material in an amount and quality which resulted in a definite increase in material learned.

I 61. Refused to explain the basis for his grading system.

I 62. Refused to admit that he was unable to answer some questions.

I 63. Read the majority of his lectures from the text or his notes.

I 64. Spoke inaudibly.

E 65. Gave unusually challenging tests which necessitated extensive preparation and thus resulted in definite learning.

E 66. Arranged for field trips, training assignments, or other experiences to enhance the meaning of the course.

E 67. Utilized audio or visual aids including blackboard illustrations to clarify lesson materials.

I 68. Hesitated or refused to see students during office hours.

I 69. Made derogatory remarks about teaching.

I 70. Deliberately wasted class time rather than covering course material.

E 71. Made a dramatic gesture to emphasize an important point.

I 72. Lectured above or below students' level of comprehension repeatedly.

E 73. Presented his lectures in the framework of an outline, either orally or on the board.

I 74. Refused to discuss returned exams.

I 75. Forced the students to shoulder the entire burden of gaining knowledge of the subject.

I 76. Course assignments remained vague and disorganized.

I 77. Lectured in a rambling, disorganized fashion.